OM of: US-09-303-518D-649 to: SPTREMBL_19:* out_format : pfs

Date: Jun 30, 2002 8:15 AM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:
-MODEL=frame+_n2p.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US09303518/runat_28062002_142714_4338/app_query.fasta_1.23501
-DB=SPTREMBL_19 -QFMT=fastan -SUFFIX=rspt -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62
-TRANS=human40.cdi -LIST=100 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=45 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09303518_@CGN1_1_1967 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-303-518D-649
Query length: 4374
Database: SPTREMBL_19:*
Database sequences: 562222
Database length: 172994929
Search time (sec): 883.700000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
sp_bacteria:Q51169	+	1615.50	1551.30	2.9e-79	1561	! Q51169 neisseria meningitidis.
sp_bacteria:Q9AIU6	+	1614.00	1549.90	3.5e-79	1552	! Q9aiu6 neisseria meningitidis.
sp_bacteriap:Q9JXK3	+	1547.00	1485.78	1.4e-75	1431	! Q9jxk3 neisseria meningitidis
sp_bacteriap:Q9JVB9	+	1518.50	1456.27	5.1e-74	1773	! Q9jvb9 neisseria meningitidis
sp_bacteriap:Q9K0B4	+	1491.50	1429.93	1.5e-72	1815	! Q9k0b4 neisseria meningitidis
sp_bacteria:Q93T34	+	1455.00	1394.86	1.3e-70	1764	! Q93t34 haemophilus aegyptius.
sp_bacteria:Q57309	+	1390.00	1337.09	3.9e-67	996	! Q57309 neisseria meningitidis.
sp_bacteria:O30575	+	1383.50	1330.79	8.8e-67	997	! O30575 neisseria meningitidis.
sp_bacteria:Q57035	+	1380.00	1327.45	1.4e-66	992	! Q57035 neisseria meningitidis.
sp_bacteria:Q9S6X4	+	1376.00	1323.58	2.2e-66	992	! Q9s6x4 neisseria meningitidis.
sp_bacteria:Q9S6X3	+	1375.50	1323.09	2.4e-66	993	! Q9s6x3 neisseria meningitidis.
sp_bacteria:Q9S6X2	+	1375.50	1323.05	2.4e-66	997	! Q9s6x2 neisseria meningitidis.
sp_bacteria:Q9S6X5	+	1375.00	1322.61	2.5e-66	992	! Q9s6x5 neisseria meningitidis.
sp_bacteria:O30573	+	1368.00	1315.84	6.0e-66	992	! O30573 neisseria meningitidis.
sp_bacteria:O30574	+	1364.00	1311.93	9.9e-66	996	! O30574 neisseria meningitidis.
sp_bacteria:Q47692	+	1111.50	1064.64	4.3e-52	1377	! Q47692 escherichia coli. tsh.
sp_bacteria:O88093	+	1107.50	1060.77	7.0e-52	1377	! O88093 escherichia coli. haemo
sp_bacteria:Q54151	+	1074.00	1028.38	4.5e-50	1372	! Q54151 shigella flexneri, and
sp_bacteria:Q9AL58	+	1074.00	1028.37	4.5e-50	1373	! Q9al58 shigella flexneri 2a. p
sp_bacteria:Q9LA54	+	1056.50	1011.69	3.9e-49	1335	! Q9la54 escherichia coli. eaac.
sp_bacteria:Q9LA58	+	1053.50	1008.78	5.7e-49	1335	! Q9la58 escherichia coli. eaaa.
sp_bacteria:Q54165	+	1001.50	958.25	3.6e-46	1366	! Q54165 shigella flexneri. sepa
sp_bacteria:Q99QC6	+	998.50	955.36	5.3e-46	1364	! Q99qc6 shigella flexneri. sepa
sp_bacteria:Q93AF7	+	975.00	932.65	9.7e-45	1359	! Q93af7 escherichia coli. autot
sp_bacteria:Q93SE3	+	941.00	899.72	6.6e-43	1363	! Q93se3 escherichia coli. hypot
sp_bacteria:Q51171	+	890.00	859.23	3.2e-40	508	! Q51171 neisseria meningitidis.
sp_bacteria:Q51165	+	883.50	853.15	7.1e-40	496	! Q51165 neisseria meningitidis.
sp_bacteria:Q51163	+	873.00	842.99	2.6e-39	496	! Q51163 neisseria meningitidis.
sp_bacteria:Q57386	+	872.00	842.02	3.0e-39	496	! Q57386 neisseria meningitidis.
sp_bacteria:Q51166	+	870.00	840.09	3.8e-39	496	! Q51166 neisseria meningitidis.
sp_bacteria:Q51167	+	870.00	839.87	3.8e-39	508	! Q51167 neisseria meningitidis.
sp_bacteria:Q51172	+	865.00	835.25	7.1e-39	496	! Q51172 neisseria meningitidis.
sp_bacteria:Q51168	+	865.00	835.03	7.1e-39	508	! Q51168 neisseria meningitidis.
sp_bacteria:Q51170	+	864.00	834.28	8.0e-39	496	! Q51170 neisseria meningitidis.
sp_bacteria:Q51164	+	856.00	826.32	2.2e-38	508	! Q51164 neisseria meningitidis.
sp_bacteria:Q51173	+	850.50	821.02	4.3e-38	507	! Q51173 neisseria meningitidis.
sp_bacteria:O69746	+	717.50	701.03	5.5e-31	192	! O69746 neisseria lactamica. iga
sp_bacteria:O69747	+	714.00	697.55	8.5e-31	194	! O69747 neisseria lactamica. iga
sp_bacteria:Q48205	+	712.50	688.29	1.2e-30	462	! Q48205 haemophilus influenzae.
sp_bacteria:Q48204	+	677.50	654.25	9.0e-29	471	! Q48204 haemophilus influenzae.
sp_bacteria:O68900	+	672.50	640.31	2.0e-28	1295	! O68900 escherichia coli. pe
sp_bacteria:O69745	+	660.50	645.96	6.5e-28	190	! O69745 neisseria lactamica.
sp_bacteria:Q48206	+	657.50	634.50	1.1e-27	492	! Q48206 haemophilus influenza
sp_bacteria:Q9EZE7	+	641.50	610.24	9.2e-27	1305	! Q9eze7 escherichia coli. en
sp_bacteria:Q9F6T1	+	636.50	605.47	1.7e-26	1295	! Q9f6t1 escherichia coli. se
sp_bacteria:O32555	+	635.00	603.99	2.1e-26	1300	! O32555 escherichia coli o15
sp_bacteria:O32591	+	635.00	603.99	2.1e-26	1300	! O32591 escherichia coli. pr
sp_bacteria:Q9L8L1	+	629.50	598.77	4.1e-26	1285	! Q91811 shigella flexneri 2a
sp_bacteria:Q48159	+	623.50	602.20	7.3e-26	460	! Q48159 haemophilus influenza
sp_bacteria:P77070	+	623.00	592.33	9.1e-26	1306	! P77070 escherichia coli. es
sp_bacteria:O69748	+	456.50	454.29	5.8e-17	100	! O69748 neisseria lactamica.
sp_bacteria:Q9XD84	+	324.50	305.93	1.1e-09	989	! Q9xd84 escherichia coli. tib
sp_bacteria:Q9RNI2	+	312.50	290.24	5.1e-09	1557	! Q9rni2 haemophilus influenz
sp_bacteriap:Q97P71	+	311.00	278.71	7.4e-09	4776	! Q97p71 streptococcus pneumo
sp_bacteria:Q9XCJ4	+	304.50	280.09	1.4e-08	2035	! Q9xcj4 salmonella typhimuri
sp_bacteria:Q9L4E2	+	297.50	280.54	3.0e-08	911	! Q914e2 bordetella bronchisep
sp_bacteria:Q9S6M9	+	290.50	273.83	7.3e-08	905	! Q9s6m9 bordetella pertussis.
sp_bacteria:Q9S3M8	+	290.50	273.83	7.3e-08	905	! Q9s3m8 bordetella pertussis.
sp_bacteria:O69259	+	288.00	271.36	9.9e-08	910	! O69259 bordetella pertussis.
sp_bacteria:O88143	+	285.50	268.89	1.4e-07	915	! O88143 bordetella pertussis.
sp_bacteriap:Q9JMS3	+	285.50	265.25	1.4e-07	1371	! Q9jms3 escherichia coli. yc
sp_bacteria:Q9KX33	+	285.50	257.14	1.7e-07	3381	! Q9kx33 streptococcus crista
sp_fungi:Q9C105	+	285.00	265.70	1.5e-07	1236	! Q9c105 schizosaccharomyces
sp_bacteria:Q9S6N0	+	284.50	268.02	1.5e-07	905	! Q9s6n0 bordetella pertussis.
sp_bacteria:Q939N5	+	284.00	256.55	2.0e-07	3072	! Q939n5 streptococcus gordon
sp_bacteria:Q93L98	+	283.00	266.42	1.8e-07	920	! Q93198 bordetella pertussis.
sp_bacteria:Q48028	+	283.00	262.16	2.0e-07	1477	! Q48028 haemophilus influenz
sp_bacteria:Q9S6N1	+	282.00	265.55	2.1e-07	910	! Q9s6n1 bordetella pertussis.
sp_bacteria:O69257	+	276.00	259.77	4.4e-07	907	! O69257 bordetella pertussis.
sp_bacteria:Q9F292	+	275.00	254.70	5.3e-07	1432	! Q9f292 yersinia pestis. yap
sp_bacteria:Q48031	+	272.50	251.65	7.4e-07	1536	! Q48031 haemophilus influenz
sp_bacteria:Q9JPJ4	+	269.50	246.26	1.1e-06	2026	! Q9jpj4 neisseria meningitid
sp_bacteriap:Q99QY4	+	265.00	240.87	2.0e-06	2271	! Q99qy4 staphylococcus aureu
sp_bacteria:Q9S6R3	+	264.00	248.78	1.9e-06	846	! Q9s6r3 escherichia coli. esp
sp_bacteria:Q9F288	+	262.00	244.72	2.6e-06	1072	! Q9f288 yersinia pestis. yap
sp_bacteriap:Q9JMS5	+	259.00	237.37	4.0e-06	1758	! Q9jms5 escherichia coli. yc
sp_bacteriap:Q9JY30	+	257.50	232.70	5.1e-06	2514	! Q9jy30 neisseria meningitid
sp_bacteria:Q9AIX8	+	255.00	240.02	5.9e-06	851	! Q9aix8 bordetella pertussis.
sp_bacteriap:Q9ZLM3	+	255.00	228.13	7.2e-06	3194	! Q9zlm3 helicobacter pylori
sp_bacteriap:Q99U54	+	255.00	221.45	8.1e-06	6713	! Q99u54 staphylococcus aureu
sp_bacteriap:Q931R6	+	255.00	221.45	8.1e-06	6713	! Q931r6 staphylococcus aureu
sp_bacteriap:Q9K0T0	+	254.00	228.66	8.0e-06	2703	! Q9k0t0 neisseria meningitid
sp_invertebrate:P91365	+	253.50	229.90	8.2e-06	2232	! P91365 caenorhabditis ele
sp_bacteria:Q9KKA4	+	253.00	232.33	8.3e-06	1615	! Q9kka4 rickettsia sp. s. om
sp_bacteria:Q9AHF9	+	253.00	226.06	9.3e-06	3241	! Q9ahf9 fusobacterium necrop
sp_bacteriap:Q9K0S7	+	252.50	230.03	9.2e-06	1975	! Q9k0s7 neisseria meningitid
sp_bacteria:Q9KKB6	+	251.00	230.40	1.1e-05	1614	! Q9kkb6 astrakhan rickettsia
sp_bacteria:Q9F289	+	250.50	230.83	1.1e-05	1457	! Q9f289 yersinia pestis. yap
sp_bacteriap:Q9ZKS9	+	250.50	226.35	1.2e-05	2399	! Q9zks9 helicobacter pylori
sp_bacteria:Q9F285	+	250.50	222.44	1.3e-05	3705	! Q9f285 yersinia pestis. yap
sp_bacteriap:O31152	+	248.50	224.90	1.5e-05	2273	! O31152 neisseria meningitid
sp_bacteriap:Q9KS12	+	248.00	218.16	1.8e-05	4558	! Q9ks12 vibrio cholerae. rtx
sp_bacteria:Q9KKA0	+	247.50	227.00	1.7e-05	1616	! Q9kka0 rickettsia honei. om
sp_bacteria:Q06454	+	247.00	239.77	1.4e-05	370	! Q06454 neisseria meningitidi
sp_bacteria:O86044	+	247.00	233.31	1.6e-05	759	! O86044 bordetella pertussis.

seq_name: sp_bacteria:Q51169

seq_documentation_block:
ID Q51169 PRELIMINARY; PRT; 1561 AA.
AC Q51169;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE IGA1 PROTEASE.
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF13;
RX MEDLINE=95302961; PubMed=7783620;

```
RA   Lomholt H., Poulsen K., Mogens K.;
RT   "Comparative characterization of the iga gene encoding IgA1 protease
RT   in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
RT   influenzae.";
RL   Mol. Microbiol. 15:495-506(1995).
DR   EMBL; X82474; CAA57857.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR002195; Dihydrooratase.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
DR   PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW   Protease.
SQ   SEQUENCE   1561 AA;  171849 MW;  1C96E291A00017D5 CRC64;


alignment_scores:
            Quality: 1615.50              Length:   1753
              Ratio:  1.766                Gaps:     55
 Percent Similarity:  52.196   Percent Identity: 28.009

alignment_block:
 US-09-303-518D-649 x Q51169   ..

 Align seg 1/1  to: Q51169  from: 1  to: 1561

      49 AAAACCGGCCGCATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTC 98
         ||||||   |||:::::::::::              |||:::|||   |||::
       2 LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18

      99 GTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGGCATCA 148
         ::::::::::   |||   :::   |||:::   :::             ::::
      18 aTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAspValA 35

     149 ACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTC 198
         ::||||||   :::||||||||||||||||||||||||||||||||   |||
      35 spTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51

     199 GGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAA 248
         ||||||   |||:::::::|||   ||||||:::|||:::   :::||||::
      52 GlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAs 68

     249 ATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAAC. 297
         ::::::::::::::::||||||||||||||||||||||:::::   |||
      68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnL 85

     298 ..GGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT 345
            :::|||::::::|||      ||||||  |||||||||  ·|||
      85 ysArgIleAlaThrValValAspProGlnTyrAlaValSerValLysHis 101

     346 .........................AACGGCGGCTATAACAACGTTGA 368
                                   |||  ||||||:::  ::
     102 AlaLysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAs 118

     369 TTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAA 418
         :     |||:::  :::|||                 |||::::
     118 nAspValAlaAspLysGluAsn.................GluTyrArgV 129

     419 TTGTGAAACGGAATAATTATAAA.................GCAGGGACT 450
         ::|||::::::|||||||||:::                |||:::::::
     129 alValGluGlnAsnAsnTyrGluProHisLysAlaTrpSerAlaSerAsn 145

     451 AAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAATT 500
            |||:::            ||||||:::|||   |||:::::::||||
     146 LeuGlyArgLeu......GluAspTyrAsnMetAlaArgPheAsnLysPh 160

     501 TGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGG...C 547
         |||||||:::::   |||:::      |||:::         |||
     160 eValThrGluValAlaProIleAlaProThrAspAlaGlyGlyGlyLeuA 177

     548 GGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCA 597
            |||   |||:::|||::::::   :::   |||||||:::||||||
     177 spThrTyrLysAspLysAsnArgPheSerSerPheValArgValGlyAla 193

     598 GGCAGGCAA......TATTGGCGATCTGATGAAGATGAGCCCAATAACCG 641
         |||||||||      |||   ::::::         :::   ::::||
     194 GlyArgGlnLeuValTyrGluLysGlyAlaTyrHisGlnGluGlyAsn.. 209

     642 CGAAAGTTCATAT.........CATATTGCAAGTGCGTATTCTTGGCTCG 682
          |||:::::::|||          :::::::::||||||   :::   :
     210 .GluLysGlyTyrAspLeuArgAspLeuSerGlnAlaTyrArgTyrAlaI 226

     683 TTGGTGGC.....................................AATACC 696
         :::::|||                                     ||||||
     226 leAlaGlyThrProTyrLysAspIleAsnIleAspGlnThrMetAsnThr 242

     697 TTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAA 746
            :::      |||   |||:::   :::::::::   ::::::|||::
     243 GluGlyLeuIleGlyPheGlyAsnHisAsnThrHisTyrSerAlaGluGl 259

     747 AATTAAACATAGC...CCATATGGTTTTTTACCAACAGGAGGCTCATTTG 793
         ::::|||::::::              |||   :::   |||   :::|
     259 uLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrGlyValLeuG 276

     794 GCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGG 843
         ||||||||||||||||||:::|||   :::|||   ||||||:::::|||
     276 lyAspSerGlySerProLeuPheAlaPheAspLysGlnLysAsnGlnTrp 292

     844 TTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAA 893
         ::::::   |||:::   :::              |||   ||||||
     293 ValPheLeuGlyThrTyrAspTyrTrpAlaGlyTyr...GlyLys..... 306

     894 TGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTG 943
                              |||:::|||          :::|||:::
     307 ..................LysSerTrpGlnGluTrpAsnIleTyr.... 315

     944 GAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCT 993
                                                    |||   :::
     316 ........................................LysLysGlu 318

     994 TTTAACGAC..............GATAATAATGGCACAGGAAAAATCAA 1028
         |||   |||                ||||||   ||||||   |||
     319 PheAlaAspLysIleLysGlnArgAspAsnAlaGlyThrIleLysGlyTy 335

    1029 TGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCG 1078
          ::::::|||:::   ::::::                    |||   |||
     335 rGlyGluHisHisTrpLysThr..................ThrGlyThr. 345

    1079 TTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTAT 1128
                 |||      :::::::::|||
     346 ...........AsnSerHisIleGlySerThr..................  352

    1129 CATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGA 1178
            |||:::      :::|||:::   |||          |||||||||::
     353 ...AlaValArgLeuAlaAsnAsnGluArgAsp...AlaAsnAsnGlyGl 367

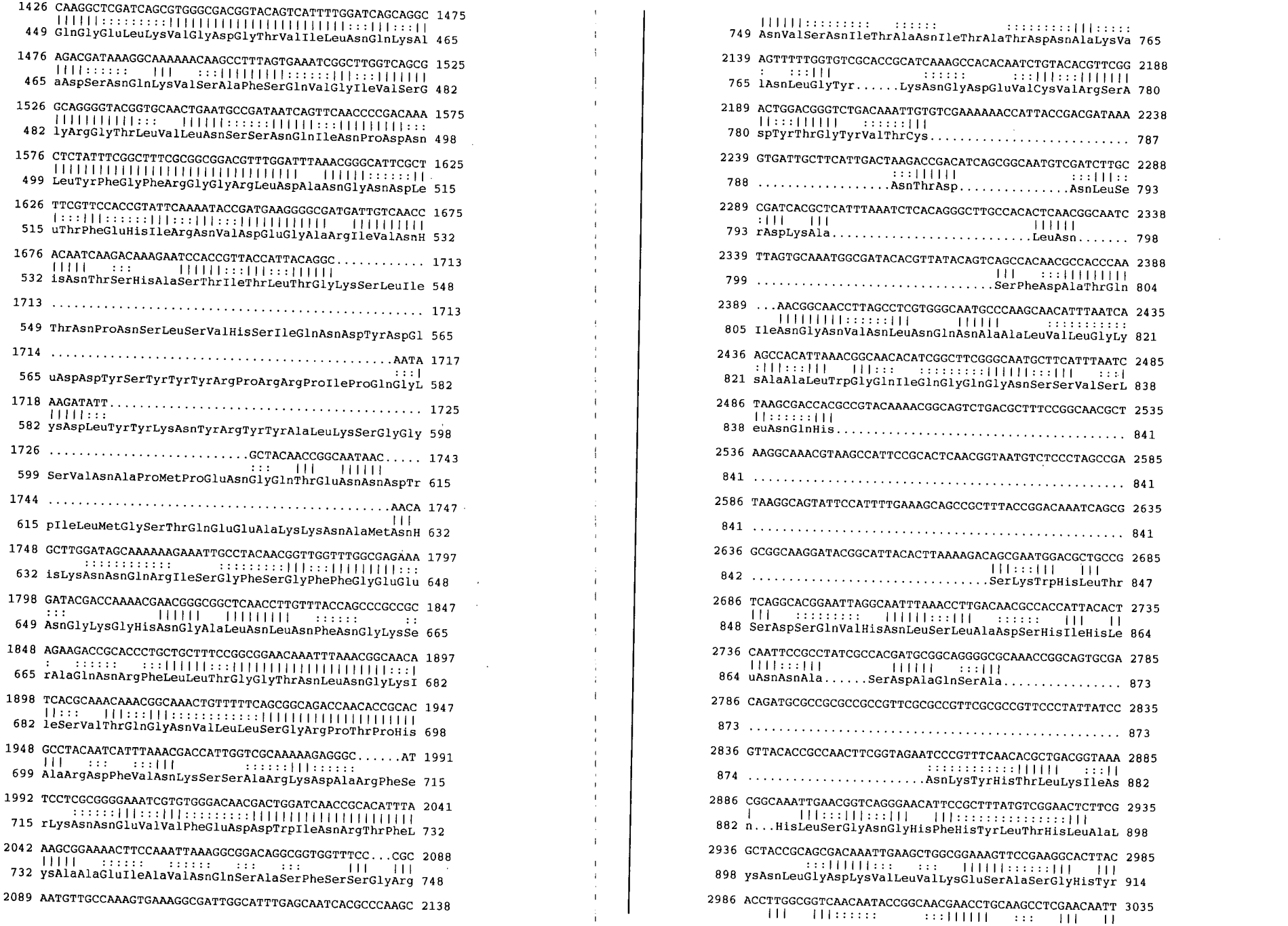
    1179 AAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCA 1228
         :|||:::::::|||       |||:::::::|||   |||:::|||   :::|
     367 nAsnValThrPhe......GluAsnAsnGlyThrLeuValLeuAspGlnA 382

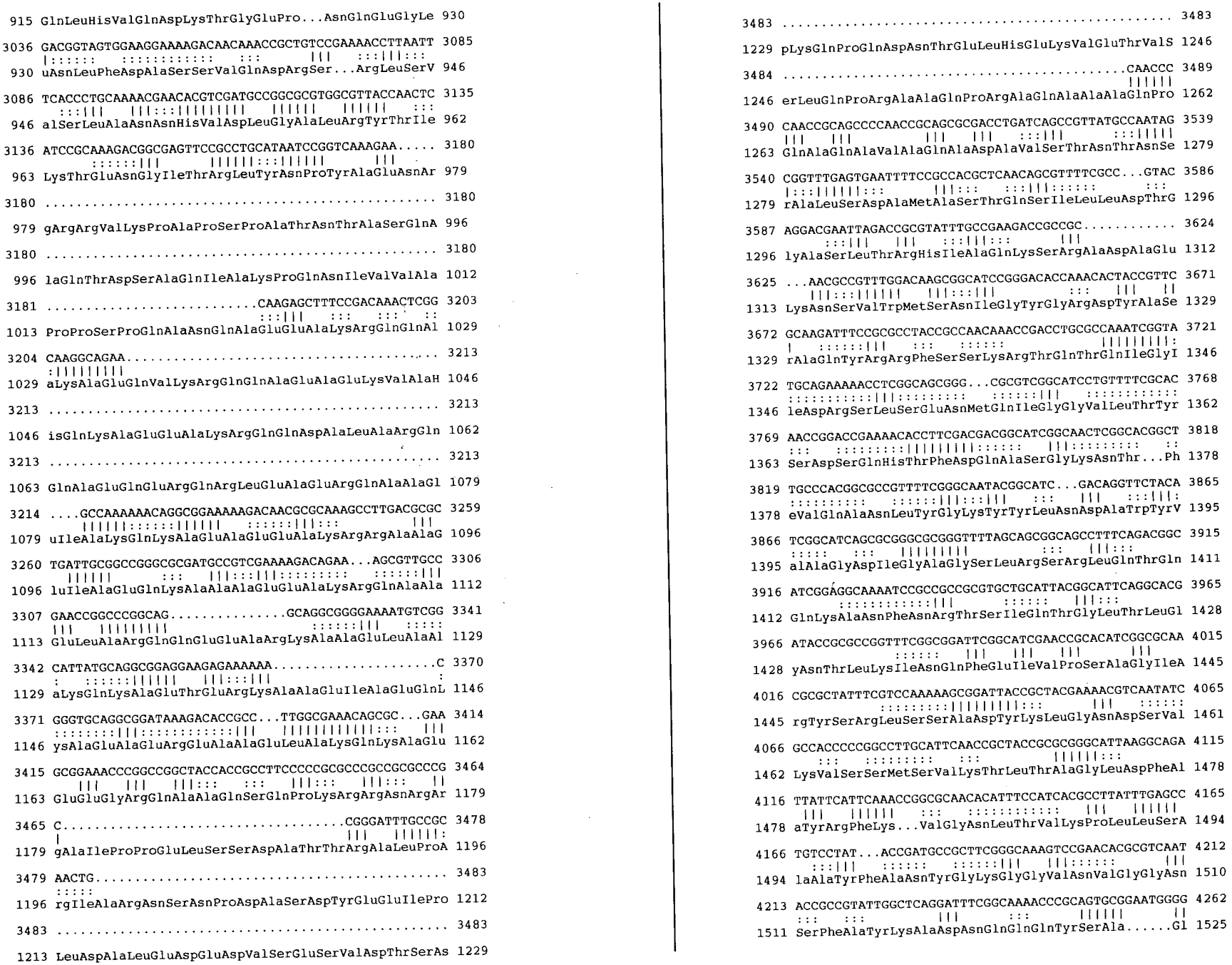
    1229 ACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTC 1278
         ||||||||||||||||||||||||||:::|||:::|||||||:::||||||
     382 snIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrThrVal 398

    1279 TCGCCTGAAAATAACGAA...ACTTGGCAAGGCGCGGGCGTTCATATCAG 1325
         :::      ||||||:::   ||||||   |||||||||:::   :::::
     399 LysGlyAlaAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspValAl 415

    1326 TGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCC 1375
         ::::   :::   |||:::|||:::|||::::::      :::|||||||
     415 aAspGlyLysLysValValTrpGlnValLysAsnProAsnGlyAspArgL 432

    1376 TGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAAC 1425
         ||:::|||||||||||||||||||||:::::::::::   |||   |||
     432 euAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyValAsn 448
```





```
 4263 CGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCG 4312
      |:::          ::::::   ||||||:::::::::::::::   :
 1525 yAlaAlaLeuLeuTyrArgAsnValThrLeuAsnValAsnGlySerIleT 1542

 4313 CCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGC 4362
      ::|||||||   |||||||||   |||   |||:::   ||||||:::
 1542 hrLysGlyLysGlnLeuGluLysGlnLysSerGlyGlnIleLysIleGln 1558

 4363 TACCGCTGG 4371
         |||:::
 1559 IleArgPhe 1561

seq_name: sp_bacteria:Q9AIU6

seq_documentation_block:
ID   Q9AIU6      PRELIMINARY;      PRT;  1552 AA.
AC   Q9AIU6;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NMB;
RA   Vitovski S., Sayers J.R.;
RT   "Degenerate specificity of Neisseria meningitidis IgA1 protease.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF235032; AAK15023.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR002195; Dihydrooratase.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
DR   PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW   Protease.
SQ   SEQUENCE   1552 AA;  171205 MW;  E49D6657B120CEFF CRC64;


alignment_scores:
          Quality: 1614.00              Length:   1739
            Ratio:  1.774                Gaps:     50
Percent Similarity:  52.329   Percent Identity: 27.430

alignment_block:
US-09-303-518D-649 x Q9AIU6   ..

Align seg 1/1  to: Q9AIU6  from: 1  to: 1552

   49 AAAACCGGCCGCATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTC 98
      ||||||  |||:::::::::::::        |||:::|||  ||||::
    2 LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18

   99 GTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGGCATCA 148
      ::::::::::  |||  :::  |||:::  :::       ::::
   18 aTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAspValA 35

  149 ACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTC 198
      ::|||||  :::|||||||||||||||||||||||||||||  |||
   35 spTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51

  199 GGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAA 248
      ||||||  |||:::::|||  ||||||:::|||:::  :::|||::
   52 GlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAs 68

  249 ATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAAC. 297
      ::::::::::::::::|||||||||||||||||||||:::::  |||
   68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnA 85

  298 ..GGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT 345
         :::  :::::::::      ||||||  |||||||||  |||

   85 rgArgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHis 101

  346 AACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGA 395
         |||  :::        :::|||  ::::::  :::  ||
  102 ValLysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAs 118

  396 T......CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATA 439
      |      :::  :::      |||:::  |||  :::|||:::|||:
  118 pValSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAspTyrG 135

  440 AAGCAGGGACTAAAGGCCATCCTTATGGCGGC.............GAT 474
      ::  :::  :::  |||  :::|||:::              |||
  135 luProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuGluAsp 150

  475 TATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGA 524
      |||:::|||  ||||||:::||||||||||||:::::  |||:::
  151 TyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaProIleAl 167

  525 AATGACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAATAATT 571
         ||||||        |||  :::  |||  |||:::|||::::
  167 aProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysAsnArgP 184

  572 ACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGAT 621
      ::  :::  ||||||:::|||||||||  |||:::
  184 heSerGluPheValArgValGlyAlaGlyThrGlnPhe..... 196

  622 GAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTA 671
         |||  |||:::|||  :::      :::::::::  |||||
  197 ......GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArgAlaTy 211

  672 TTCTTGGCTCGTTGGTGGCAATACCTTT.............GCACAAA 706
      |   :::  ::::::|||:::  :::                ::::::|
  211 rArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValThrSerA 228

  707 ATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACAT 756
      ||  :::      |||  :::::::::|||:::::::::      |||
  228 snLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLysHisHis 244

  757 AGCCCA.....................TATGGTTTTTTACCAACAGG 782
      ||||||                        :::  |||  :::
  245 SerProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuThrAsnTy 261

  783 AGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAA 832
      :::  :::|||||||||||||||||:::|||  ||||||  |||:
  261 rAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrAspLysGlnG 278

  833 AGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATA 882
      :::::::::|||::::::  |||:::  :::          |||
  278 luLysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGlyTyr... 293

  883 GGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG...TTCTATGA 929
         ::::::|||:::::::|||           :::|||  :::|||
  294 ...GlnLysAsnSerTrpGln.....GluTrpAsnIleTyrLy 305

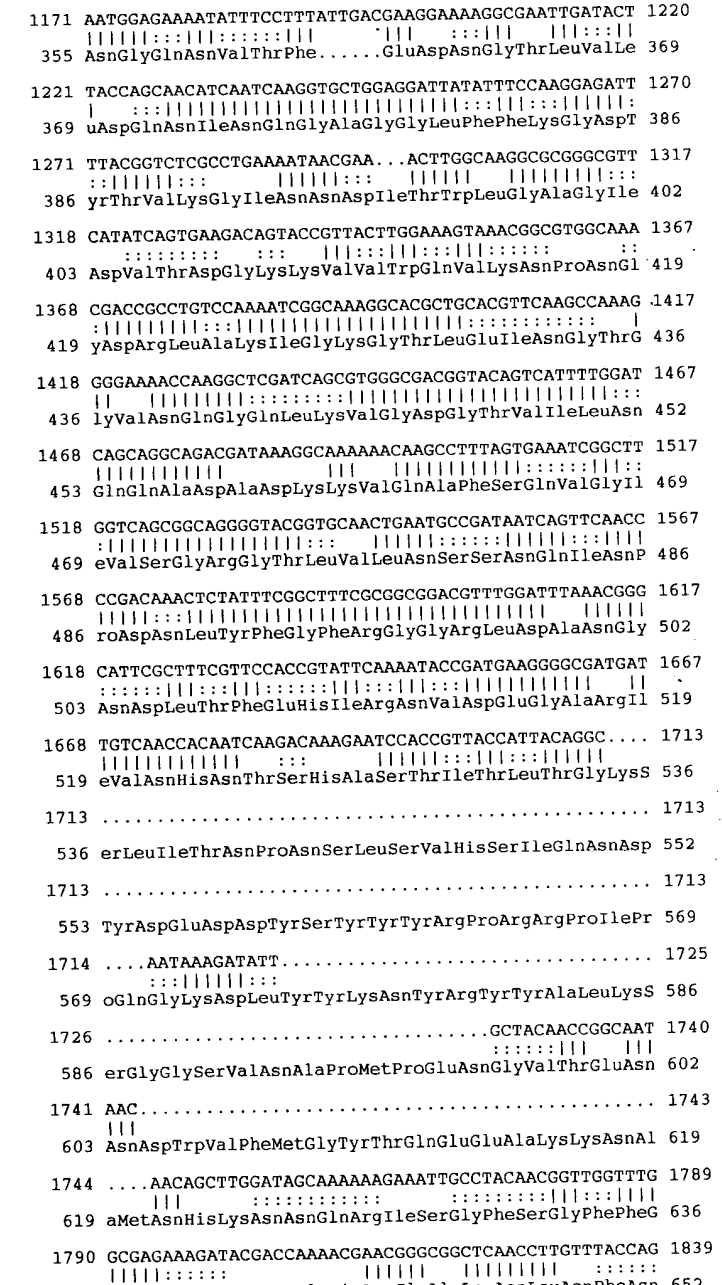
  930 TGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAA 979
      :::  ||||||                    :::  ||||||:
  305 sLysGluPheAlaAsp.....LysIleArgGlnA 315

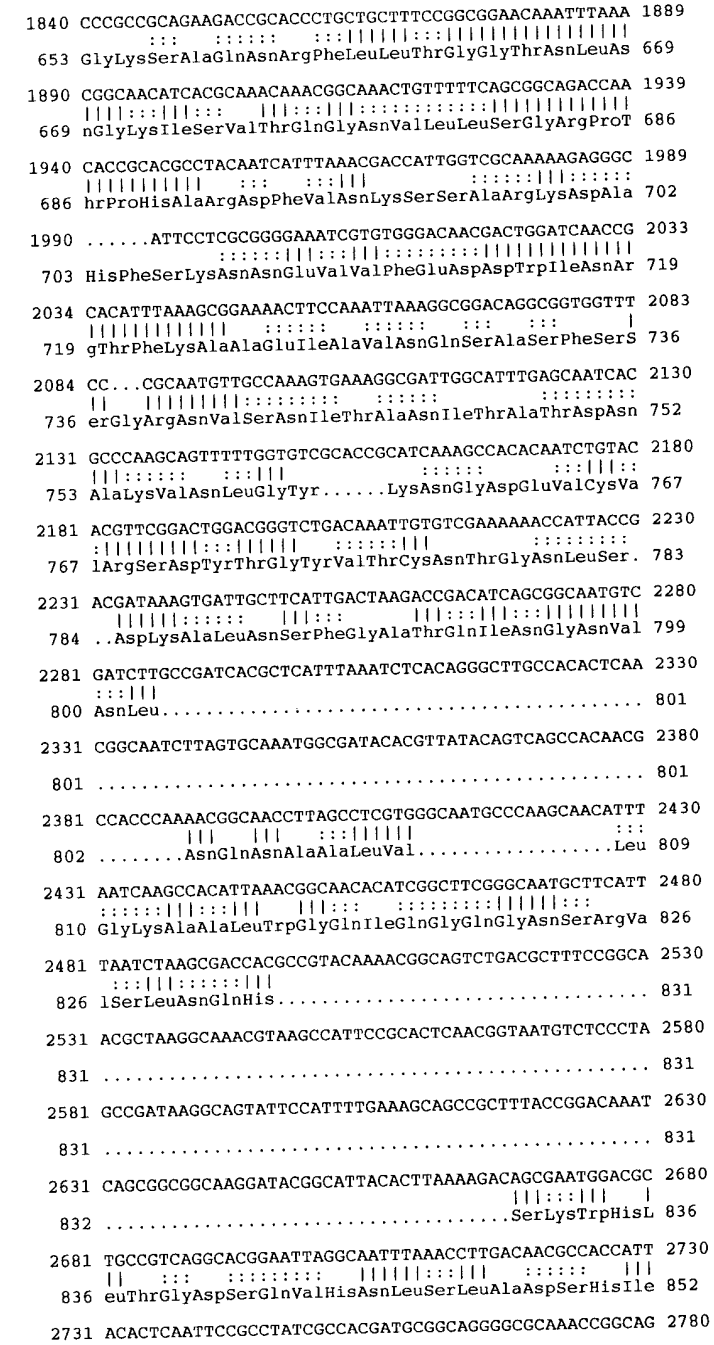
  980 ATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAATCAAT 1029
      ::                  ||||||  ||||||  |||  |||
  315 rg.....AspAsnAlaGlyThrIleLysGlyAsn 324

 1030 GCCAAACATGAACACAAT.........TCTCTGCCTAATAGATTAAAAAC 1070
      ::::::|||:::  |||       :::      :::::::::  ::
  325 GlyGluHisHisTrpAsnIleThrPheGlyThrAsnSerLysIleGlySe 341

 1071 ACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAAC 1120
      :  :::|||:::|||
  341 rThrAlaValArgLeu.............................. 346

 1121 CTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAAT 1170
                 ||||||:::  ::::::              |||
  347 .............AlaGlyAsnGluLysAsp.....AlaAsn 354
```


1171 AATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACT 1220
355 AsnGlyGlnAsnValThrPhe.....GluAspAsnGlyThrLeuValLe 369
1221 TACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATT 1270
369 uAspGlnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspT 386
1271 TTACGGTCTCGCCTGAAAATAACGAA...ACTTGGCAAGGCGCGGGCGTT 1317
386 yrThrValLysGlyIleAsnAsnAspIleThrTrpLeuGlyAlaGlyIle 402
1318 CATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAA 1367
403 AspValThrAspGlyLysLysValValTrpGlnValLysAsnProAsnGl 419
1368 CGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAG 1417
419 yAspArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrG 436
1418 GGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGAT 1467
436 lyValAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsn 452
1468 CAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTT 1517
453 GlnGlnAlaAspAlaAspLysLysValGlnAlaPheSerGlnValGlyIl 469
1518 GGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACC 1567
469 eValSerGlyArgGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnP 486
1568 CCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGG 1617
486 roAspAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGly 502
1618 CATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGAT 1667
503 AsnAspLeuThrPheGluHisIleArgAsnValAspGluGlyAlaArgIl 519
1668 TGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGC.... 1713
519 eValAsnHisAsnThrSerHisAlaSerThrIleThrLeuThrGlyLysS 536
1713 1713
536 erLeuIleThrAsnProAsnSerLeuSerValHisSerIleGlnAsnAsp 552
1713 1713
553 TyrAspGluAspAspTyrSerTyrTyrTyrArgProArgArgProIlePr 569
1714 ....AATAAAGATATT..... 1725
569 oGlnGlyLysAspLeuTyrTyrLysAsnTyrArgTyrTyrAlaLeuLysS 586
1726 .....GCTACAACCGGCAAT 1740
586 erGlyGlySerValAsnAlaProMetProGluAsnGlyValThrGluAsn 602
1741 AAC..... 1743
603 AsnAspTrpValPheMetGlyTyrThrGlnGluGluAlaLysLysAsnAl 619
1744 ....AACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTG 1789
619 aMetAsnHisLysAsnAsnGlnArgIleSerGlyPheSerGlyPhePheG 636
1790 GCGAGAAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAG 1839
636 lyGluGluAsnGlyLysGlyHisAsnGlyAlaLeuAsnLeuAsnPheAsn 652


1840 CCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAA 1889
653 GlyLysSerAlaGlnAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAs 669
1890 CGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAA 1939
669 nGlyLysIleSerValThrGlnGlyAsnValLeuLeuSerGlyArgProT 686
1940 CACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGC 1989
686 hrProHisAlaArgAspPheValAsnLysSerSerAlaArgLysAspAla 702
1990 ......ATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCG 2033
703 HisPheSerLysAsnAsnGluValValPheGluAspAspTrpIleAsnAr 719
2034 CACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTT 2083
719 gThrPheLysAlaAlaGluIleAlaValAsnGlnSerAlaSerPheSerS 736
2084 CC...CGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCAC 2130
736 erGlyArgAsnValSerAsnIleThrAlaAsnIleThrAlaThrAspAsn 752
2131 GCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTAC 2180
753 AlaLysValAsnLeuGlyTyr......LysAsnGlyAspGluValCysVa 767
2181 ACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACCG 2230
767 lArgSerAspTyrThrGlyTyrValThrCysAsnThrGlyAsnLeuSer. 783
2231 ACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTC 2280
784 ..AspLysAlaLeuAsnSerPheGlyAlaThrGlnIleAsnGlyAsnVal 799
2281 GATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAA 2330
800 AsnLeu..... 801
2331 CGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACG 2380
801 801
2381 CCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTT 2430
802 .....AsnGlnAsnAlaAlaLeuVal.....Leu 809
2431 AATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCTTCATT 2480
810 GlyLysAlaAlaLeuTrpGlyGlnIleGlnGlyGlnGlyAsnSerArgVa 826
2481 TAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCA 2530
826 lSerLeuAsnGlnHis..... 831
2531 ACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTA 2580
831 831
2581 GCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAAT 2630
831 831
2631 CAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGC 2680
832 .....SerLysTrpHisL 836
2681 TGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATT 2730
836 euThrGlyAspSerGlnValHisAsnLeuSerLeuAlaAspSerHisIle 852
2731 ACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAAACCGGCAG 2780

```
                 |||||||:::|||          ||||||   :::|||
  853 HisLeuAsnAsnAla......SerAspAlaGlnSerAla..........  863

 2781 TGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTAT 2830

  863 .................................................. 863

 2831 TATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACG 2880
                                  :::::::::::::||||||
  864 ............................AsnGlnTyrHisThrLeuLys 870

 2881 GTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACT 2930
      :::|||      |||:::|||:::|||   |||:::::::::::::::||
  871 IleAsn...HisLeuSerGlyAsnGlyHisPheHisTyrLeuThrHisLe 886

 2931 CTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCA 2980
      |            :::|||||||:::   :::   ||||||:::::|||
  886 uAlaGluAsnLeuGlyAspLysValLeuValLysGluSerAlaSerGlyH 903

 2981 CTTACACCTTGGCGGTCAACAATACCGGCAACGAACCTGCAAGCCTCGAA 3030
        |||   |||   |||::::::      :::||||||   :::   |||
  903 isTyrGlnLeuHisValGlnAspLysThrGlyGluPro...AsnGlnGlu 918

 3031 CAATTGACGGTAGTGGAAGGAAAAGACAACAAACCGCTGTCCGAAAACCT 3080
        |||::::::   ::::::::::::   :::      |||   :::||
  919 GlyLeuAsnLeuPheAspAlaSerSerValGlnAspArgSer...ArgLe 934

 3081 TAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACC 3130
      |:::   :::|||   |||:::|||||||||   ||||||   ||||||
  934 uSerValSerLeuAlaAsnAsnHisValAspLeuGlyAlaLeuArgTyrT 951

 3131 AACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAA 3180
        :::      :::::|||      ||||||:::||||||     |||
  951 hrIleLysThrGluAsnGlyIleThrArgLeuTyrAsnProTyrAlaGlu 967

 3180 .................................................. 3180

  968 AsnArgArgArgValLysProValProSerProAlaThrAsnThrAlaSe 984

 3180 .................................................. 3180

  984 rGlnAlaGlnThrAspSerAlaGlnIleAlaLysProGlnAsnIleValV 1001

 3181 ...............................CAAGAGCTTTCCGACAAA 3198
                                     ::::||   :::   :::
 1001 alAlaProProSerProGlnAlaAsnGlnAlaGluGluAlaLysArgGln 1017

 3199 CTCGGCAAGGCAGAAGCCAAAAAA.............CAGGCGGAAAAAGA 3236
        :::||||||||||      |||            :::|||||:::
 1018 GlnAlaLysAlaGluGlnValLysArgGlnGlnAlaGluAlaGluArgLy 1034

 3237 CAACGCGCAA.......................AGCCTTGACGCGCTGA 3262
        :::|||:::                     :::   :::|||
 1034 sSerAlaGluLeuAlaLysGlnLysAlaGluAlaGluArgGluAlaArgG 1051

 3263 TTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCG 3312
          |||   |||:::   :::|||:::   :::|||   ||||||
 1051 luLeuAlaThrArgGlnLysAlaGluGlnGluArgSerSerAlaGluLeu 1067

 3313 GCCCGG.................................................. 3318
      ||||||
 1068 AlaArgArgHisGluLysGluArgGluAlaAlaGluLeuSerAlaLysGl 1084

 3318 .................................................. 3318

 1084 nLysValGluAlaGluArgGluAlaGlnAlaLeuAlaValArgArgLysA 1101

 3318 .................................................. 3318

 1101 laGluAlaGluGluAlaLysArgGlnAlaAlaGluLeuAlaArgGlnGln 1117

 3319 .........CAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGA 3359
                ::   :::::::|||   ::::::   ::::::||||||
```

```
 1118 GluGluAlaArgLysAlaAlaGluLeuAlaAlaLysGlnLysAlaGluTh 1134

 3360 AGAGAAAAAA.....................CGGGTGCAGGCGGATAAAG 3388
        |||:::|||                    ::::::::::|||:::::::
 1134 rGluArgLysAlaAlaGluIleAlaGluGlnLysAlaGluAlaGluArgG 1151

 3389 ACACCGCC...TTGGCGAAACAGCGC...GAAGCGGAAACCCGGCCGGCT 3432
      :::::|||   ||||||||||||:::   |||   |||   |||   |||
 1151 luAlaAlaGluLeuAlaLysGlnLysAlaGluGluGluGlyArgGlnAla 1167

 3433 ACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGC..................... 3465
      :::   :::   |||:::   |||:::   |||
 1168 AlaGlnSerGlnProLysArgArgAsnArgArgAlaIleProProGluLe 1184

 3466 ..................CGGGATTTGCCGCAACTG............... 3483
                          |||   ||||||::::::
 1184 uSerSerAspAlaThrThrArgAlaLeuProArgIleAlaArgAsnSerA 1201

 3483 .................................................. 3483

 1201 snProAspAlaSerAspTyrGluLysIleProLeuAspAlaLeuGluAsp 1217

 3483 .................................................. 3483

 1218 GluAspValSerGluSerValAspThrSerAspLysGlnProGlnAspAs 1234

 3483 .................................................. 3483

 1234 nThrGluLeuHisGluLysValGluThrValSerLeuGlnProArgAlaA 1251

 3484 ..CAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTAT 3531
        ||||||:::   |||         |||   |||   :::|||
 1251 laGlnProArgAlaGlnAlaAlaAlaGlnAlaAspAlaValSerThrAsn 1267

 3532 GCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGC 3581
      :::||||||:::||||||||:::      |||:::   :::|||::::::
 1268 ThrAsnSerAlaLeuSerAspAlaMetAlaSerThrGlnSerIleLeuLe 1284

 3582 C...GTACAGGACGAATTAGACCGCGTATTTGCCGAAGACCGCCGC.... 3624
        :::      :::|||   |||   :::|||:::      |||
 1284 uAspThrGlyAlaSerLeuThrArgHisIleAlaGlnLysSerArgAlaA 1301

 3625 ..........AACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACAC 3663
              |||:::||||||   |||:::|||          :::
 1301 spAlaGluLysAsnSerValTrpMetSerAsnIleGlyTyrGlyArgAsp 1317

 3664 TACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCA 3713
      |||   |||   ::::::|||   :::   ::::::          ||
 1318 TyrAlaSerAlaGlnTyrArgArgPheSerSerLysArgThrGlnThrGl 1334

 3714 AATCGGTATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGT 3760
      |||||||:::::::::::::|||:::::::::   :::::::|||   ::::
 1334 nIleGlyIleAspArgSerLeuSerGluAsnMetGlnIleGlyGlyValL 1351

 3761 TTTCGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCAACTCG 3810
      ::::::::::::   :::::::::|||||||||::::::   |||::::::
 1351 euThrTyrSerAspSerGlnHisThrPheAspGlnAlaSerGlyLysAsn 1367

 3811 GCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATC...GACAG 3857
      :::   ::::::::::::   ::::::|||:::|||   :::   |||
 1368 Thr...PheValGlnAlaAsnLeuTyrGlyLysTyrTyrLeuAsnAspAl 1383

 3858 GTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTT 3907
        :::|||::::::   :::   |||||||||          :::   |||:
 1383 aTrpTyrValAlaGlyAspIleGlyAlaGlySerLeuArgSerArgLeuG 1400

 3908 CAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATT 3957
      ::            ::::::::::::|||      ::::::   |||:::
 1400 lnThrGlnGlnLysAlaAsnPheAsnArgThrSerIleGlnThrGlyLeu 1416

 3958 CAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACAT 4007
                              ::::::   |||   |||   |||
 1417 ThrLeuGlyAsnThrLeuLysIleAsnGlnPheGluIleValProSerAl 1433
```

```
  4008 CGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACG 4057
       |||               ::::::::::||||||||||:::      |||
  1433 aGlyIleArgTyrSerArgLeuSerSerAlaAspTyrLysLeuGlyAsnA 1450

  4058 TCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATT 4107
         ::::::   :::   :::::::::   :::         ||||||:::
  1450 spSerValLysValSerSerMetSerValLysThrLeuThrAlaGlyLeu 1466

  4108 AAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTA 4157
              |||   ||||||   :::   ::::::::::::   |||
  1467 AspPheAlaTyrArgPheLys...ValGlyAsnLeuThrValLysProLe 1482

  4158 TTTGAGCCTGTCCTAT...ACCGATGCCGCTTCGGGCAAAGTCCGAACAC 4204
        ||||||   :::|||   ::::::   :::::::|||   |||::::::
  1482 uLeuSerAlaAlaTyrPheAlaAsnTyrGlyLysGlyGlyValAsnValG 1499

  4205 GCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCG 4254
             |||:::   :::         |||      :::      ||||||
  1499 lyGlyAsnSerPheAlaTyrLysAlaAspAsnGlnGlnGlnTyrSerAla 1515

  4255 GAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGC 4304
             |||:::            ::::::   ||||||:::::::::::
  1516 ......GlyAlaAlaLeuLeuTyrArgAsnValThrLeuAsnValAsnGl 1530

  4305 TGCCGCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCA 4354
       ::::   :::||||||   |||||||||   |||   ||||:::   ||||
  1530 ySerIleThrLysGlyLysGlnLeuGluLysGlnLysSerGlyGlnIleL 1547

  4355 AATTAGGCTACCGCTGG 4371
       ||:::      |||:::
  1547 ysIleGlnIleArgPhe 1552

seq_name: sp_bacteriap:Q9JXK3

seq_documentation_block:
ID   Q9JXK3      PRELIMINARY;      PRT;  1431 AA.
AC   Q9JXK3;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SERINE-TYPE PEPTIDASE.
GN   NMB1998.
OS   Neisseria meningitidis (serogroup B).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / SEROGROUP B;
RX   MEDLINE=20175755; PubMed=10710307;
RA   Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA   Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA   Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA   Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA   Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA   Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA   Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA   Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT   "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT   MC58.";
RL   Science 287:1809-1815(2000).
DR   EMBL; AE002549; AAF42325.1; -.
DR   MEROPS; S06.UPW; -.
DR   TIGR; NMB1998; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
KW   Complete proteome.
SQ   SEQUENCE   1431 AA;  157645 MW;  6801F0E209C24EFC CRC64;

alignment_scores:
            Quality: 1547.00            Length:   1669
                Ratio:  1.809                 Gaps:    50
  Percent Similarity:  51.228   Percent Identity: 27.741

alignment_block:
 US-09-303-518D-649 x Q9JXK3   ..

 Align seg 1/1  to: Q9JXK3  from: 1  to: 1431

      55 GGCCGCATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGG 104
         |||   :::      ||||||||||||   :::|||      :::
      17 GlyLeuPheAlaValSerProAlaTyrSerSerIleValArgAsn..... 31

     105 CATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGGCATCAACTACC 154
                                                 ::::::||||
      32 .....AspValAspTyrG 36

     155 AATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCG 204
         |||||:::|||||||||||||||||||||||||   |||:::|||||||||
      36 lnTyrPheArgAspPheAlaGluAsnLysGlyAlaPheThrValGlyAla 52

     205 AAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAAT 254
         :::::::|||::::::   :::|||:::|||:::::::::|||:::   ::
      53 SerAsnIleSerIleGlnAspLysGlnGlyLysIleLeuGlyArgValLe 69

     255 GACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGT...AACGGCG 301
         ::::      ||||||   ||||||   |||   :::|||   :::::::
      69 uAsnGlyIleProMetProAspPheArgValSerAsnArgGlnThrAlaI 86

     302 TGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGC 351
         ::|||:::||||||      ||||||:::   ||||||   ||||||
      86 leAlaThrLeuValHisProGlnTyrValAsnSerValLysHisAsnVal 102

     352 GGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACA 401
         ||||||:::::::::::||||||   :::   :::||||||:::::::::
     103 GlyTyrGlySerIleGlnPheGlyAsnAspThrGlnAsnProGluGluGl 119

     402 TCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCAGGGACTA 451
         :   :::||||||:::::|||:::||||||
     119 nAlaTyrThrTyrArgLeuValSerArgAsn..................... 129

     452 AAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAATTT 501
               ||||||      ||||||||:::||||||||||:::|||:::
     130 ..ProHisPro...AspTyrAspTyrHisLeuProArgLeuAsnLysLeu 144

     502 GTCACAGATGCAGAACCTGTTGAAATGACCAGT..... 534
         ||||||:::   :::|||:::   ::::::|||
     145 ValThrGluIleSerProThrAlaLeuSerSerValProLeuLeuGlyAs 161

     535 ....TATATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACC 580
                    ::::::   |||:::|||   ::::::::::|||
     161 nGlyGlnProLysAlaAsnAlaTyrLeuAspThrAspArgPheProTyrP 178

     581 GTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAG 630
           ||||||:::|||:::|||   |||      |||:::   :::
     178 heValArgLeuGlySerGlyThrGlnGlnValArgLysAlaAspGly... 193

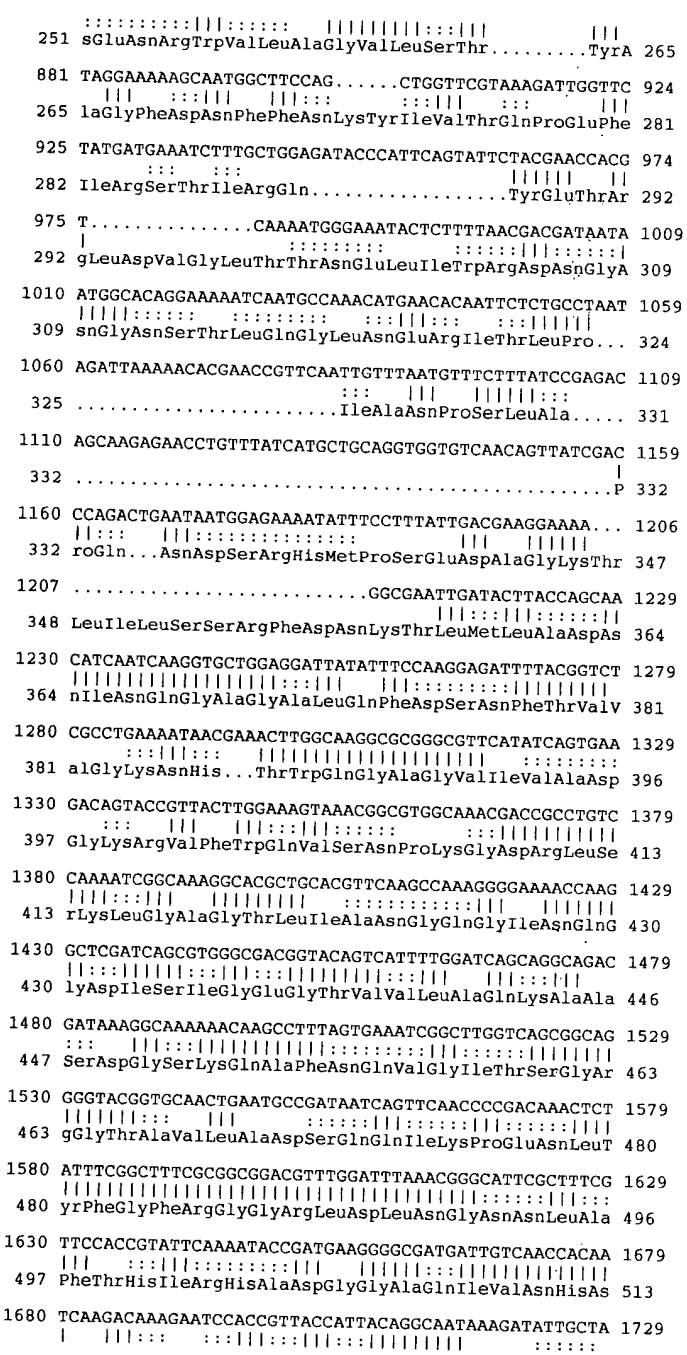
     631 CCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCT 680
                       :::      :::   |||   ||||||:::::||
     194 ..............ThrArgThrArgThrAlaProAlaTyrGlnTyrLe 205

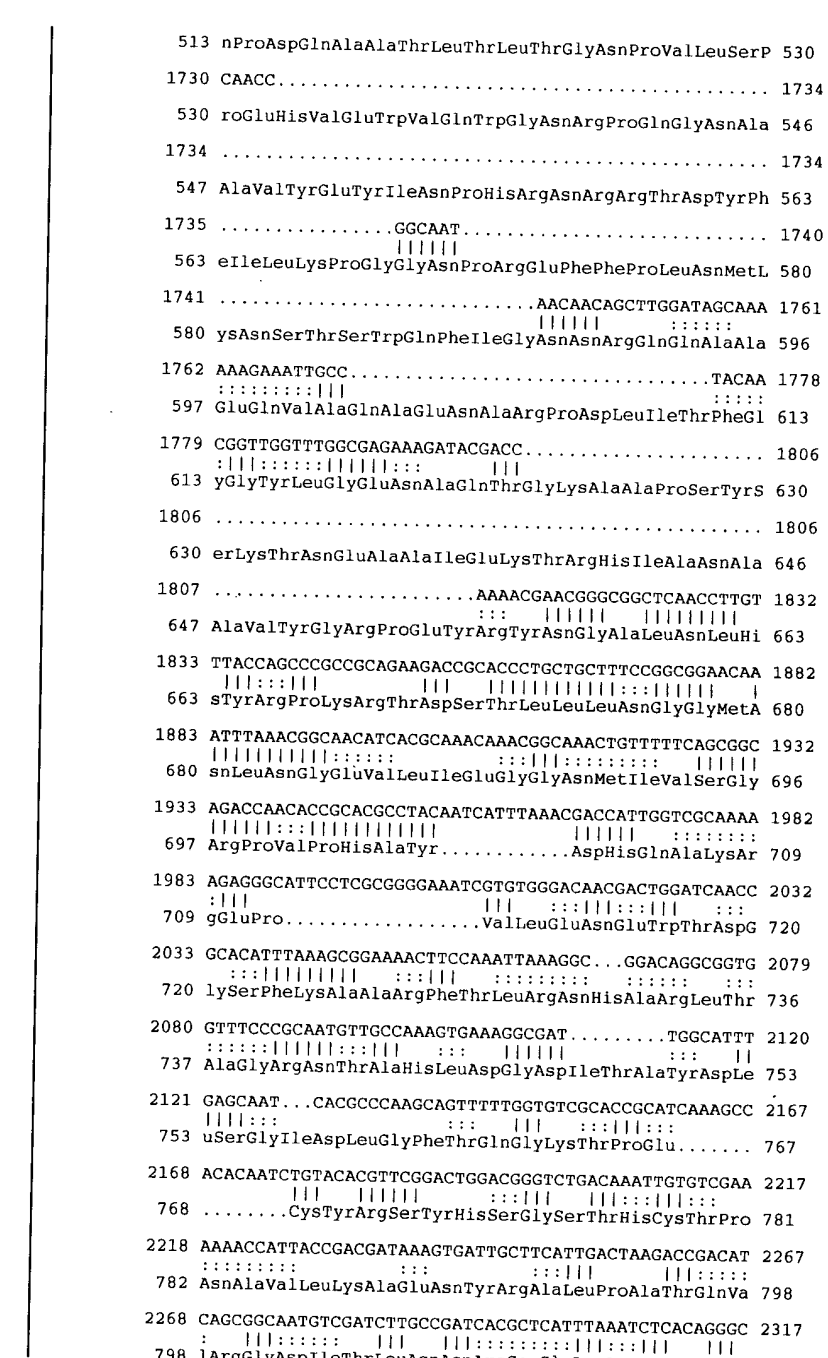
     681 CGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCA 730
         |:::||||||:::   :::         |||      :::      :::
     205 uThrGlyGlyThrProLeuLysValLeuGlyPheGlnAsnHisGlyLeuL 222

     731 ACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACA 780
           :::|||:::   ::::::      :::|||         |||   |||
     222 euValGlyGly...SerLeuThrAspGlnPro.........LeuAsnThr 234

     781 GGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCA 830
           :::      ||||||||||||||||||:::|||   :::|||   ::
     235 TyrAlaIleAlaGlyAspSerGlySerProLeuPheAlaPheAspLysHi 251

     831 AAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATA 880
```


251 sGluAsnArgTrpValLeuAlaGlyValLeuSerThr.....TyrA 265
881 TAGGAAAAAGCAATGGCTTCCAG.....CTGGTTCGTAAAGATTGGTTC 924
265 laGlyPheAspAsnPhePheAsnLysTyrIleValThrGlnProGluPhe 281
925 TATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACG 974
282 IleArgSerThrIleArgGln.....TyrGluThrAr 292
975 T.....CAAAATGGGAAATACTCTTTTAACGACGATAATA 1009
292 gLeuAspValGlyLeuThrThrAsnGluLeuIleTrpArgAspAsnGlyA 309
1010 ATGGCACAGGAAAAATCAATGCCAAACATGAACACAATTCTCTGCCTAAT 1059
309 snGlyAsnSerThrLeuGlnGlyLeuAsnGluArgIleThrLeuPro... 324
1060 AGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGAC 1109
325 .....IleAlaAsnProSerLeuAla..... 331
1110 AGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGAC 1159
332 .....P 332
1160 CCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAA... 1206
332 roGln...AsnAspSerArgHisMetProSerGluAspAlaGlyLysThr 347
1207 .....GGCGAATTGATACTTACCAGCAA 1229
348 LeuIleLeuSerSerArgPheAspAsnLysThrLeuMetLeuAlaAspAs 364
1230 CATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCT 1279
364 nIleAsnGlnGlyAlaGlyAlaLeuGlnPheAspSerAsnPheThrValV 381
1280 CGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAA 1329
381 alGlyLysAsnHis...ThrTrpGlnGlyAlaGlyValIleValAlaAsp 396
1330 GACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTC 1379
397 GlyLysArgValPheTrpGlnValSerAsnProLysGlyAspArgLeuSe 413
1380 CAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAG 1429
413 rLysLeuGlyAlaGlyThrLeuIleAlaAsnGlyGlnGlyIleAsnGlnG 430
1430 GCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGAC 1479
430 lyAspIleSerIleGlyGluGlyThrValValLeuAlaGlnLysAlaAla 446
1480 GATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAG 1529
447 SerAspGlySerLysGlnAlaPheAsnGlnValGlyIleThrSerGlyAr 463
1530 GGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCT 1579
463 gGlyThrAlaValLeuAlaAspSerGlnGlnIleLysProGluAsnLeuT 480
1580 ATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCG 1629
480 yrPheGlyPheArgGlyGlyArgLeuAspLeuAsnGlyAsnAsnLeuAla 496
1630 TTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAA 1679
497 PheThrHisIleArgHisAlaAspGlyGlyAlaGlnIleValAsnHisAs 513
1680 TCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTA 1729


513 nProAspGlnAlaAlaThrLeuThrLeuThrGlyAsnProValLeuSerP 530
1730 CAACC..... 1734
530 roGluHisValGluTrpValGlnTrpGlyAsnArgProGlnGlyAsnAla 546
1734 1734
547 AlaValTyrGluTyrIleAsnProHisArgAsnArgArgThrAspTyrPh 563
1735 .....GGCAAT..... 1740
563 eIleLeuLysProGlyGlyAsnProArgGluPhePheProLeuAsnMetL 580
1741 .....AACAACAGCTTGGATAGCAAA 1761
580 ysAsnSerThrSerTrpGlnPheIleGlyAsnAsnArgGlnGlnAlaAla 596
1762 AAAGAAATTGCC.....TACAA 1778
597 GluGlnValAlaGlnAlaGluAsnAlaArgProAspLeuIleThrPheGl 613
1779 CGGTTGGTTTGGCGAGAAAGATACGACC..... 1806
613 yGlyTyrLeuGlyGluAsnAlaGlnThrGlyLysAlaAlaProSerTyrS 630
1806 1806
630 erLysThrAsnGluAlaAlaIleGluLysThrArgHisIleAlaAsnAla 646
1807 .....AAAACGAACGGGCGGCTCAACCTTGT 1832
647 AlaValTyrGlyArgProGluTyrArgTyrAsnGlyAlaLeuAsnLeuHi 663
1833 TTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAA 1882
663 sTyrArgProLysArgThrAspSerThrLeuLeuLeuAsnGlyGlyMetA 680
1883 ATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGC 1932
680 snLeuAsnGlyGluValLeuIleGluGlyGlyAsnMetIleValSerGly 696
1933 AGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAA 1982
697 ArgProValProHisAlaTyr.....AspHisGlnAlaLysAr 709
1983 AGAGGGCATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACC 2032
709 gGluPro.....ValLeuGluAsnGluTrpThrAspG 720
2033 GCACATTTAAAGCGGAAAACTTCCAAATTAAAGGC...GGACAGGCGGTG 2079
720 lySerPheLysAlaAlaArgPheThrLeuArgAsnHisAlaArgLeuThr 736
2080 GTTTCCCGCAATGTTGCCAAAGTGAAAGGCGAT.....TGGCATTT 2120
737 AlaGlyArgAsnThrAlaHisLeuAspGlyAspIleThrAlaTyrAspLe 753
2121 GAGCAAT...CACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCC 2167
753 uSerGlyIleAspLeuGlyPheThrGlnGlyLysThrProGlu..... 767
2168 ACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAA 2217
768 .....CysTyrArgSerTyrHisSerGlySerThrHisCysThrPro 781
2218 AAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACAT 2267
782 AsnAlaValLeuLysAlaGluAsnTyrArgAlaLeuProAlaThrGlnVa 798
2268 CAGCGGCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGC 2317
798 lArgGlyAspIleThrLeuAsnAspArgSerGluLeuArgLeu...GlyL 814

```
2318 TTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACA 2367
      |||   |||   |||::::::   |||
 814 ysAlaHisLeuTyrGlySerIleArgAla..................... 823

2368 GTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGC 2417

 823 ..................................................... 823

2418 CCAAGCAACATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGG 2467

 823 ..................................................... 823

2468 GCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTG 2517

 823 ..................................................... 823

2518 ACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGG 2567

 823 ..................................................... 823

2568 TAATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCT 2617

 823 ..................................................... 823

2618 TTACCGGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAA... 2664
                          ||||||||||||||||::::::::::::
 824 ...................GlyLysAspThrAlaValArgMetGluAla 833

2665 GACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCT 2714
     ||||||:::|||||||||   ::::::::::   |||   |||:::||
 834 AspSerAsnTrpThrLeuSerGlnSerSerHisThrGlyAlaLeuThrLe 850

2715 TGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAG 2764
     ||||:::|||   ||||||||||||      :::   ::::::
 850 uAspGlyAlaGlnIleThrLeuAsnProAspPheAlaAsnAsn..... 864

2765 GGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGT 2814

 864 ..................................................... 864

2815 TCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCG 2864
                                        :::   :::::::||
 865 ....................................ThrHisAsnAsnAr 869

2865 TTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCC 2914
     |||||||||||||||||||||||||   |||:::|||   ||||||||||
 869 gPheAsnThrLeuThrValAsnGlyThrLeuAspGlyPheGlyThrPheA 886

2915 GCTTTATGTCGGAACTCTTC...GGCTACCGCAGCGACAAATTGAAGCTG 2961
     |||||::::::   :::          ::::::     |||||||||
 886 rgPheLeuThrGlyIleValArgLysGlnAsnAlaProProLeuLysLeu 902

2962 GCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACCGGCAA 3011
          :::|||:::|||::::::   :::   |||:::||||||||::
 903 GluGlyAspSerArgGlyAlaPheGlnIleHisValLysAsnThrGlyGl 919

3012 CGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACA 3061
     :||||||   :::   |||:::|||::::::|||      :::   |||
 919 nGluProGlnThrThrGluSerLeuAlaLeuVal......SerLeuAsnP 934

3062 AACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGAT 3111
               |||::::::   :::|||||||||||||||   :::::::|||
 934 roLysHisSerHisGlnAlaArgPheThrLeuGlnAsnGlyTyrAlaAsp 950

3112 GCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCT 3161
        ||||||:::||||||   |||   ::::::::::::   :::   ||
 951 LeuGlyAlaTyrArgTyrIleLeuArgLysAsnAsnAsnGlyTyrSerLe 967

3162 GCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAG 3211
     |:::||||||:::||||||   |||||||:::   :::      :::||||
 967 uTyrAsnProLeuLysGluAlaGluLeuGlnIleGluAlaThrArgAlaG 984
```

```
3212 AA.....GCC 3216
     ||
 984 luHisGluArgAsnGlnGlnAlaTyrAsnGlnLeuGlnAlaThrAspIle 1000

3217 AAAAAACAGGCGGAAAAAGACAAC..... 3240
     ::::::|||::::::   |||:::
1001 SerArgGlnValGlnHisAspSerAspAlaThrArgGlnAlaLeuGlnAl 1017

3240 ..... 3240

1017 aTrpGlnAsnSerGlnThrGluLeuAlaArgIleAspSerGlnValGlnT 1034

3240 ..... 3240

1034 yrLeuSerAlaGlnLeuLysGlnThrAspProLeuThrGlyIleLeuThr 1050

3241 ...GCGCAAAGCCTTGACGCGCTG..... 3261
        |||||||:::|||   |||
1051 ArgAlaGlnAsnLeuCysAlaAlaGlnGlyTyrSerAlaAspIleCysAr 1067

3262 ....ATTGCGGCCGGGCGCGATGCC..... 3282
         :::|||   :::   |||:::
1067 gGlnValAlaLysAlaAlaAspThrAsnAspLeuThrLeuPheGluThrG 1084

3283 .....GTCGAAAAGACAGAAAGCGTTGCCGAA 3309
                    ||||||   :::|||||||      :::
1084 luLeuAspThrTyrIleGluArgValGluMetAlaGluSerGluLeuAsp 1100

3310 CCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGA 3359
        ||||||||||:::|||                  ||||||   ||
1101 LysAlaArgGlnGlyGlyAsp..............AlaGlnAlaValGl 1112

3360 AGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAG. 3408
     |      :::::::::      :::      :::   |||::::::|||
1112 uThrAlaArgHisAlaTyrLeuAsnAlaLeuAsnArgLeuSerArgGlnI 1129

3409 ..CGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGC 3456
       ::::::   :::|||      |||         :::|||:::
1129 leHisSerLeuLysThrGlyValAlaGlyIleArgMetProAsnLeu... 1144

3457 CGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACC 3506

1144 ..... 1144

3507 GCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTT 3556
           :::|||||||||||||   ||||||:::::::::||||||   :
1145 ....AlaGluLeuIleSerArgSerAlaAsnThrAlaValSerGluGlnA 1160

3557 CCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTA 3606
     ::|||      ::::::         :::      ::::::||||||
1160 laAlaTyrAsnThrGlyArgGlnGlnAlaGlyArgArgIleAspArgHis 1176

3607 TTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACAC 3656
     :::::::::   :::::::::   :::|||   :::         :::
1177 LeuThrAspProGlnGlnGlnAsnIleTrpLeuGluThrGlyThrGlnGl 1193

3657 CAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACC 3706
           |||:::|||       |||   |||:::|||   |||:::
1193 nThrAspTyrHisSerGlyThrHisArgProTyrGlnGlnThrThrAsnT 1210

3707 TGCGCCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGC 3753
        :::||||||:::|||   ::::::   :::         ||||||
1210 yrAlaHisIleGlyIleGlnThrGlyIleThrAspArgLeuSerValGly 1226

3754 ATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGG 3803
        :::::::::   :::|||||||:::|||   ||||||:::|||:::::
1227 ThrIleLeuThrAspGluArgThrAsnAsnArgPheAspGluGlyValSe 1243

3804 C......AACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACG 3847
     :      |||   ::::::   ||||||      |||   |||:::   |
1243 rAlaArgAsnArgSerAsnGlyAlaHisLeuPheValLysGlyGluAsnG 1260

3848 GCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGC 3897
```

```
      ||                :::      :::|||      |||:::|||:::
 1260 lyAla.............LeuPheAlaAlaAlaAspLeuGlyTyrSerAsn 1272

 3898 GGC......AGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGT 3941
      :::          ::::::|||      |||:::   :::||||||:::::
 1273 SerArgThrArgPheThrAspTyrAspGlyAlaAlaValArgArgHisAl 1289

 3942 GCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCG 3991
      :          ||||||:::::      :::             |||:::
 1289 aTrpAspAlaGlyIleAsnThrGlyIleLysIleAsp...ThrGlyIleA 1305

 3992 GCATCGAACCGCACATCGGC.....................GCAACGCGC 4020
      :::::::::|||:::   |||                     :::::::|||
 1305 snLeuArgProTyrAlaGlyIleArgIleAsnArgSerAsnGlyAsnArg 1321

 4021 TATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCAC 4070
      |||   ::::::   |||:::                        |||   ::
 1322 TyrValLeuAspGlyAlaGlu.....................IleAsnSe 1331

 4071 CCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATT 4120
      :|||:::          :::   ::::::||||||||||:::   |||   :
 1331 rProAlaGlnIleGlnThrThrTrpHisAlaGlyIleArgLeuAspLysT 1348

 4121 CATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCC 4170
      ::   :::   :::|||      ::::::||||||   :::|||   :::
 1348 hrValGluLeuGlyGln...AlaLysLeuThrProAlaPheSerSerAsp 1363

 4171 TAT......ACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATAC 4214
      |||       |||         ||||||:::::   :::   ||||||
 1364 TyrTyrHisThrArgGlnAsnSerGlySerAlaLeuSer...ValAsnAs 1379

 4215 CGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCG 4264
      :::|||   |||:::   :::                          :
 1379 pArgThrLeuLeuGlnGlnAlaAlaHisGlyThrLeuHisThrLeuGlnI 1396

 4265 TAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCC 4314
      :::::|||      ||||||:::::   :::||||||||||||   :::
 1396 leAspAlaGlyTyrLysGlyTrpAsnAlaLysLeuHisAlaAlaTyrGly 1412

 4315 AAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTA 4364
      |||      :::         :::   :::|||||||||||||:::|||||
 1413 LysAspSerAsnThrAlaArgHisLysGlnAlaGlyIleLysIleGlyTy 1429

 4365 CCGCTGG 4371
      |:::|||
 1429 rAsnTrp 1431
```

seq_name: sp_bacteriap:Q9JVB9

seq_documentation_block:

```
ID   Q9JVB9      PRELIMINARY;      PRT;  1773 AA.
AC   Q9JVB9;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (EC 3.4.21.7).
GN   IGA OR NMA0905.
OS   Neisseria meningitidis (serogroup A).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=65699;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX   MEDLINE=20222556; PubMed=10761919;
RA   Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA   Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA   Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA   Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA   Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA   Whitehead S., Spratt B.G., Barrell B.G.;
RT   "Complete DNA sequence of a serogroup A strain of Neisseria
RT   meningitidis Z2491.";
RL   Nature 404:502-506(2000).
DR   EMBL; AL162754; CAB84182.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR002195; Dihydrooratase.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
DR   PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW   Protease; Hydrolase; Complete proteome.
SQ   SEQUENCE   1773 AA;  196350 MW;  CAC19E7313D76CE1 CRC64;
```

alignment_scores:

```
           Quality: 1519.50              Length:   1836
             Ratio:  1.562                Gaps:     57
 Percent Similarity:  52.996   Percent Identity: 27.397
```

alignment_block:

US-09-303-518D-649 x Q9JVB9 ..

Align seg 1/1 to: Q9JVB9 from: 1 to: 1773

```
      49 AAAACCGGCCGCATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTC 98
         ||||||   |||:::::::::::::         |||:::|||   |||::
       2 LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18

      99 GTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGGCATCA 148
         ::::::::::   |||   :::   |||:::   :::             ::::
      18 aTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAspValA 35

     149 ACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTC 198
         ::||||||   :::||||||||||||||||||||||||||||||||   |||
      35 spTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51

     199 GGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAA 248
         ||||||   |||:::::|||   ||||||:::|||:::   :::|||::
      52 GlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAs 68

     249 ATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAAC. 297
         ::::::::::::::::|||||||||||||||||||||:::::   |||
      68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnA 85

     298 ..GGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT 345
            :::   :::::::::      ||||||   ||||||||   |||
      85 rgArgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHis 101

     346 AACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGA 395
            |||   :::         :::|||   ::::::   :::   ||
     102 ValLysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAs 118

     396 T......CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATA 439
         |       :::   :::      |||:::   |||   :::|||:::|||:
     118 pValSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAspTyrG 135

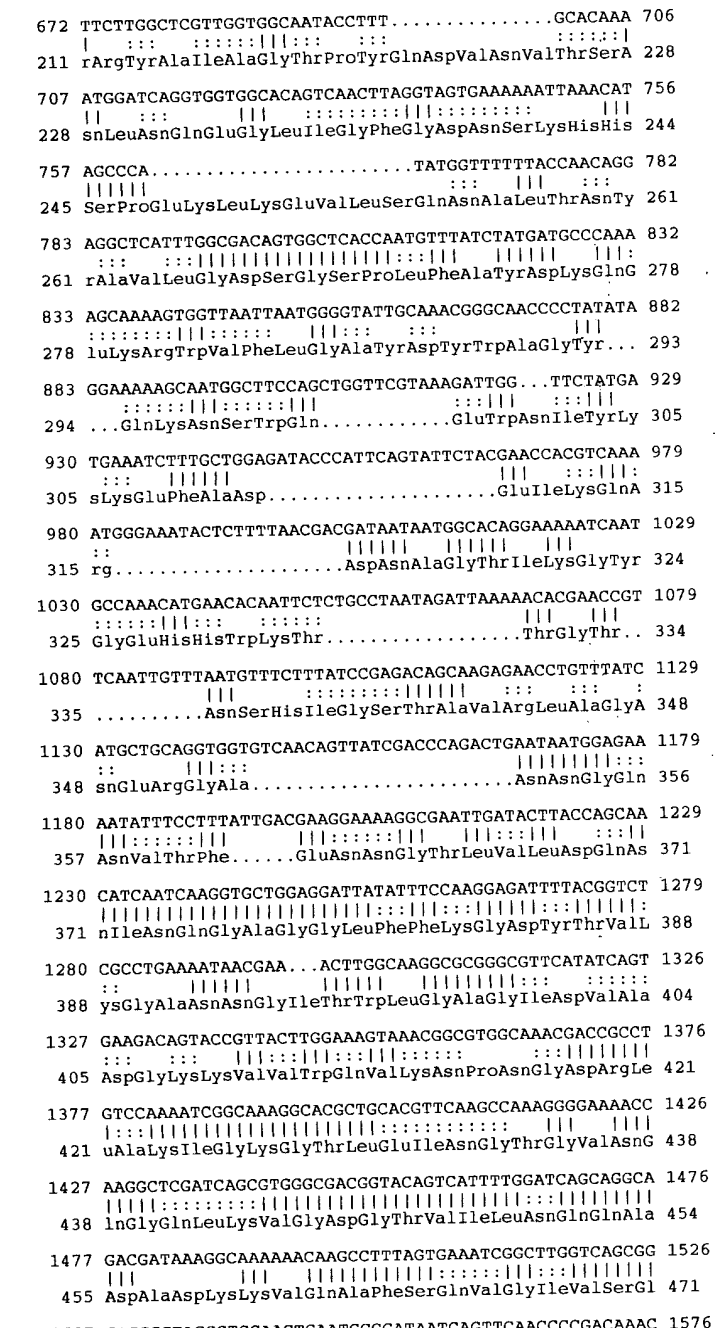
     440 AAGCAGGGACTAAAGGCCATCCTTATGGCGGC.............GAT 474
         ::   :::   :::   |||   :::|||:::                |||
     135 luProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuGluAsp 150

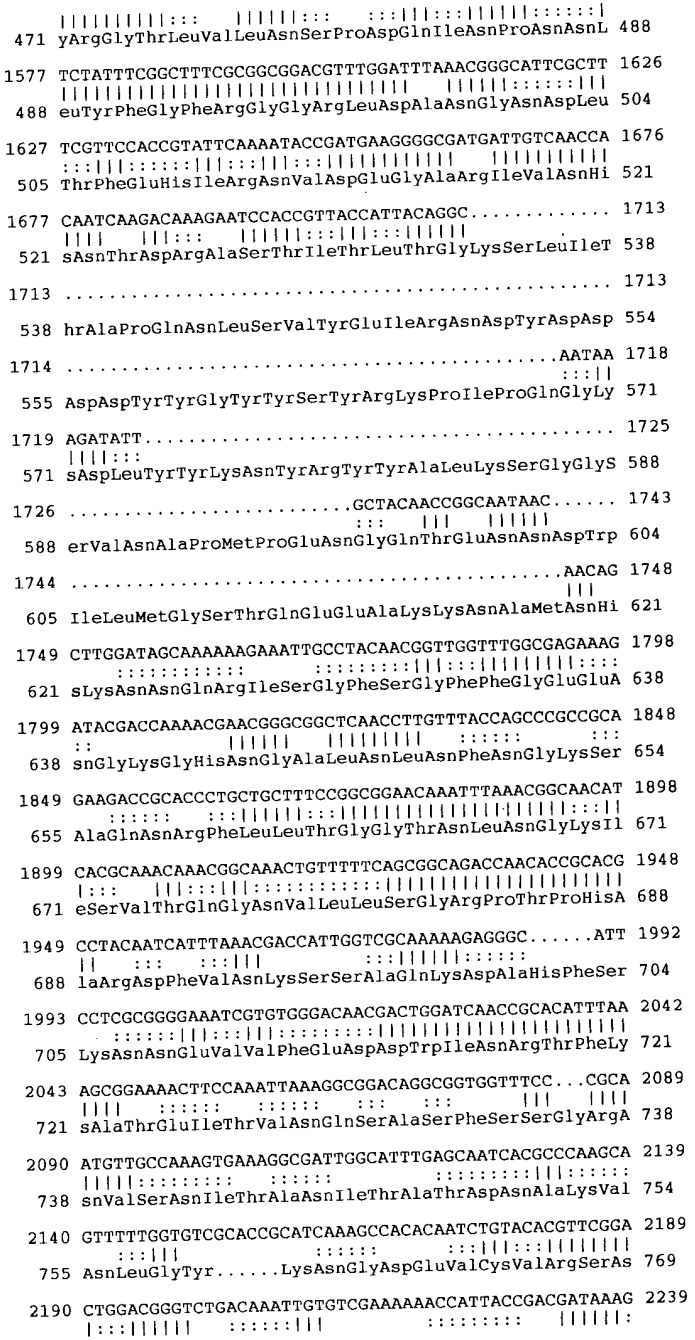
     475 TATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGA 524
         |||:::|||   |||||||:::||||||||||||:::::   |||:::
     151 TyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaProIleAl 167

     525 AATGACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAATAATT 571
            ||||||         |||   :::   |||   |||:::|||::::
     167 aProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysAsnArgP 184

     572 ACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGAT 621
         ::   :::   ||||||:::|||||||||   |||:::
     184 heSerGluPheValArgValGlyAlaGlyThrGlnPhe............ 196

     622 GAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTA 671
            |||   |||:::|||   :::      :::::::::   |||||
     197 ......GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArgAlaTy 211
```






769 pTyrThrGlyTyrValThrCysAsnThrGlyAsnLeuSer...AspLysA 785
2240 TGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTCGATCTTGCC 2289
785 laLeuAsnSerPheGlyAlaThrLysIleAsnGlyAsnValAsnLeuAsn 801
2290 GATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCT 2339
802 GlnAsnAlaAlaLeuValLeu...GlyLysAlaAlaLeuTrpGlyGlnIl 817
2340 TAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCC..... 2382
817 eGlnGlyGlnGlyAsnSerArgValSerLeuAsnGlnHisSerLysTrpH 834
2383 .....ACCCAAAACGGCAACCTTAGCCTC..... 2406
834 isLeuThrGlyAspSerGlnValHisAsnLeuSerLeuAlaAspSerHis 850
2407 .....GTGGGCAATGCCCAAGCAACATTTAATCAAGCCAC 2441
851 IleHisLeuAsnAsnAlaSerAspAlaGlnSerAlaAsnGlnTyrHisTh 867
2442 ATTAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCG 2491
867 rLeuLysIleAsn...HisLeuSerGlyAsnGlyHisPheHisTyrLeuT 883
2492 ACCACGCCGTACAAAAC...GGCAGTCTGACGCTT..... 2523
883 hrHisLeuAlaGluAsnLeuGlyAspLysValLeuValLysGluSerAla 899
2524 TCCGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAAT.. 2571
900 SerGlyHisTyrGlnLeuHisValGlnAspLysThrGlyGluProAsnGl 916
2571 2571
916 nGluGlyLeuAsnLeuPheAspAlaSerSerValArgAspArgSerHisL 933
2572 .....GTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGC 2616
933 euSerValSerLeuAlaAsnHisHisVal...AspLeuGlyAlaLeuArg 948
2617 TTTACCGGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGA 2666
949 TyrThrIleLysThrGluAsnGly...IleThrArgLeuTyrAsnProTy 964
2667 CAGCGAA.....TGGACGCTGCCGTCAGGCACGG 2695
964 rAlaGluAsnArgArgArgValLysProAlaProSerProAlaThrAsnT 981
2696 AATTAGGCAATTTAAACCTTGACAACGCCACCATTACA..... 2733
981 hrAlaSerGlnAlaGlnThrAspSerAlaGlnIleAlaLysProGlnAsn 997
2734 .....CTCAATTCCGCCTA 2747
998 IleValValAlaProProSerProGlnAlaAsnGlnAlaGluGluAlaLy 1014
2748 TCGCCACGATGCGGCAGGGGCGCAAACC..... 2775
1014 sArgGlnGlnAlaLysAlaGluGlnValLysArgGlnGlnAlaGluAlaG 1031
2775 2775
1031 luArgLysSerAlaGluLeuAlaLysGlnLysAlaGluAlaGluArgGlu 1047
2776 .....GGCAGTGCGAC 2786
1048 AlaArgGluLeuAlaThrArgGlnLysAlaGluGlnGluArgSerSerAl 1064
2787 AGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCG 2836
1064 aGluLeuAlaArgArgHisGluLysGluArgGluAlaAlaGluLeuSerA 1081
2837 TTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAAC 2886
1081 laLysGlnLysValGluAlaGluArgGluAlaGlnAlaLeuAlaValArg 1097
2887 GGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTT.CG 2935
1098 ArgLysAlaGluAlaGluGluAlaLysArgGlnAlaAlaGluLeuAlaAr 1114
2936 GCTACCGCAGCGACAAATTGAAGCT..... 2960
1114 gArgHisGluLysGluArgGluAlaAlaGluLeuSerAlaLysGlnArgV 1131
2961 ..GGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACCGG 3008
1131 alGlyGluGluGluArgArgGlnThrAlaGlnSerGlnProGlnArgArg 1147
3009 CAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGG..... 3050
1148 LysArgArgAlaAlaProGlnAspTyrMetAlaAlaSerGlnAspArgPr 1164
3050 3050
1164 oLysArgArgGlyHisArgSerValGlnGlnAsnAsnValGluIleAlaG 1181
3051 .....AAAAGACAACAA..... 3062
1181 lnAlaGlnAlaGluLeuAlaArgArgGlnGlnGluGluArgLysAlaAla 1197
3062 3062
1198 GluLeuLeuAlaLysGlnArgAlaGluAlaGluArgGluAlaGlnAlaLe 1214
3063 .ACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGAT 3111
1214 uAlaAlaArgArgLysAlaGluAlaGluGluAlaLysArg.GlnAlaAla 1230
3112 GCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCT 3161
1231 GluLeuAlaHisArgGlnGluAlaGluArgLysAlaAlaGluLeuSerAl 1247
3162 GCATAAT......CCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCA 3205
1247 aAsnGlnLysAlaAlaAlaGluAlaGlnAlaLeuAlaAlaArgGlnGlnL 1264
3206 AGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTT... 3252
1264 ysAlaLeuAlaArgGlnGlnGluGluAlaArgLysAlaAlaGluLeuAla 1280
3253 .....GACGCGCTGATTGCGGCCGG 3272
1281 ValLysGlnLysAlaGluThrGluArgLysThrAlaGluLeuAlaLysGl 1297
3273 GCGCGATGCCGTCGAA.....AAGACAGAAAGCGTTG 3304
1297 nArgAlaAlaAlaGluAlaAlaLysArgGlnGlnGluAlaArgGlnThrA 1314
3305 CCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCG 3354
1314 laGluLeuAlaArgArgGlnGluAlaGluArg.....GlnAla 1326
3355 GAGGAA......GAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTT 3398
1327 AlaGluLeuSerAlaLysGlnLysAlaGluThrAspArgGluAlaAlaGl 1343
3399 GGCGAAACAGCGCGAAGCGGAAACC.....CGGCCGGCTACCACCG 3439
1343 uSerAlaLysArgLysAlaGluGluGluGluHisArgGlnAlaAlaGlnS 1360
3440 CCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAA..... 3480
1360 erGlnProGln...ArgArgLysArgArgAlaAlaProGlnAspTyrMet 1375

```
3480 ..................................................  3480

1376 AlaAlaSerGlnAsnArgProLysArgArgGlyArgArgSerThrLeuPr 1392

3480 ..................................................  3480

1392 oAlaProProSerProSerPheAspSerSerAlaTyrAlaAlaProArgA 1409

3480 ..................................................  3480

1409 laLeuHisAsnProAspTrpTyrGluAsnAspTyrGluGluIleProLeu 1425

3480 ..................................................  3480

1426 AspAlaLeuGluAspGluAsnValSerGluSerValAspThrSerAspLy 1442

3480 ..................................................  3480

1442 sGlnProGlnAspAsnThrGluLeuHisGluLysValGluThrValSerL 1459

3481 ..................................CTGCAACCCCAA  3492
                                                  |||||||||
1459 euGlnProArgAlaAlaGlnProArgAlaGlnAlaAlaThrGlnProGln 1475

3493 CCGCAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGG  3542
       |||          |||   |||   :::|||      :::|||||||::
1476 AlaGlnAlaAlaAlaGlnAlaAspAlaValSerThrAsnThrAsnSerAl 1492

3543 TTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCC...GTACAGG  3589
     :|||||||:::      |||:::   :::|||:::::      :::
1492 aLeuSerAspAlaMetAlaSerThrGlnSerIleLeuLeuAspThrGlyA 1509

3590 ACGAATTAGACCGCGTATTTGCCGAAGACCGCCGC.................  3624
       :::|||   |||   :::|||:::      |||
1509 laSerLeuThrArgHisIleAlaGlnLysSerArgAlaAspAlaGluLys 1525

3625 AACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCA  3674
     |||:::||||||   |||:::|||         :::   |||   |||
1526 AsnSerValTrpMetSerAsnIleGlyTyrGlyArgAspTyrAlaSerAl 1542

3675 AGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGC  3724
      :::::::|||   :::   ::::::          |||||||||::::
1542 aGlnTyrArgArgPheSerSerLysArgThrGlnThrGlnIleGlyIleA 1559

3725 AGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTTCGCACAAC  3771
     ::::::::|||:::::::::   :::::|||   ::::::::::::::::
1559 spArgSerLeuSerGluAsnMetGlnIleGlyGlyValLeuThrTyrSer 1575

3772 CGGACCGAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGC  3821
       :::::::::|||||||||::::::   |||:::::::::   :::::
1576 AspSerGlnHisThrPheAspGlnAlaSerGlyLysAsnThr...PheVa 1591

3822 CCACGGCGCCGTTTTCGGGCAATACGGCATC...GACAGGTTCTACATCG  3868
     :::::::   ::::::|||:::|||   :::   |||   :::|||::::
1591 lGlnAlaAsnLeuTyrGlyLysTyrTyrLeuAsnAspAlaTrpTyrMetA 1608

3869 GCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATC  3918
     ::   :::   |||||||||         :::   |||:::
1608 laGlyAspIleGlyAlaGlySerLeuArgSerArgLeuGlnThrGlnGln 1624

3919 GGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATA  3968
       :::::::::::|||      ::::::   |||:::
1625 LysAlaAsnPheAsnArgThrSerIleGlnThrGlyLeuThrLeuGlyAs 1641

3969 CCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGC  4018
               ::::::   |||   |||   |||      |||
1641 nThrLeuLysIleAsnGlnPheGluIleValProSerAlaGlyIleArgT 1658

4019 GCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCC  4068
             :::::::::|||||||||:::      |||   ::::::
1658 yrSerArgLeuSerSerAlaAspTyrLysLeuGlyAsnAspSerValLys 1674

4069 ACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTA  4118
       :::   :::::::::   :::         ||||||:::         ||
1675 ValSerSerMetSerValLysThrLeuThrAlaGlyLeuAspPheAlaTy 1691

4119 TTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGT  4168
       |   ||||||   :::   ::::::::::::   |||   ||||||   :
1691 rArgPheLys...ValGlyAsnLeuThrValLysProLeuLeuSerAlaA 1707

4169 CCTAT...ACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACC  4215
     ::|||   ::::::   ::::::|||   |||::::::      |||:::
1707 laTyrPheAlaAsnTyrGlyLysGlyGlyValAsnValGlyGlyAsnSer 1723

4216 GCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGT  4265
       |||         |||      :::      ||||||      |||::
1724 PheValTyrLysAlaAspAsnGlnGlnGlnTyrSerAla......GlyAl 1738

4266 AAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCA  4315
     :                ::::::   |||||||:::::::::::::::   :::|
1738 aAlaLeuLeuTyrArgAsnValThrLeuAsnValAsnGlySerIleThrL 1755

4316 AAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTAC  4365
     |||||   |||||||||   |||   |||:::   |||||||:::
1755 ysGlyLysGlnLeuGluLysGlnLysSerGlyGlnIleLysIleGlnIle 1771

4366 CGCTGG 4371
     |||:::
1772 ArgPhe 1773
```

```
seq_name: sp_bacteriap:Q9K0B4

seq_documentation_block:
ID   Q9K0B4      PRELIMINARY;      PRT;  1815 AA.
AC   Q9K0B4;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA-SPECIFIC SERINE ENDOPEPTIDASE.
GN   NMB0700.
OS   Neisseria meningitidis (serogroup B).
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=491;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=MC58 / SEROGROUP B;
RX   MEDLINE=20175755; PubMed=10710307;
RA   Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA   Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA   Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA   Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA   Mason T., Ciecko A., Parksey D.S., Blair E., Cittone H., Clark E.B.,
RA   Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,
RA   Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,
RA   Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RT   "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT   MC58.";
RL   Science 287:1809-1815(2000).
DR   EMBL; AE002424; AAF41117.1; -.
DR   MEROPS; S06.001; -.
DR   TIGR; NMB0700; -.
DR   InterPro; IPR002195; Dihydrooratase.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
DR   PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW   Complete proteome.
SQ   SEQUENCE   1815 AA;  201077 MW;  2259D4D71762C57F CRC64;


alignment_scores:
            Quality: 1491.50              Length:   2000
              Ratio:  1.637                Gaps:     56
 Percent Similarity:  45.550   Percent Identity: 24.550

alignment_block:
```

US-09-303-518D-649 x Q9K0B4 ..

Align seq 1/1 to: Q9K0B4 from: 1 to: 1815

```
 49 AAAACCGGCCGCATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTC 98
    ||||||   |||:::::::::::::        ||||:|||   ||||::
  2 LysThrLysArgPheLysIleAsnAlaIleSerLeuSerIlePheLeuAl 18

 99 GTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGGCATCA 148
    ::::::::::   |||   :::   |||:::   :::         ::::
 18 aTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAspValA 35

149 ACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTC 198
    ::||||||   :::||||||||||||||||||||||||||||||||   |||
 35 spTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPhePheVal 51

199 GGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAA 248
    ||||||   |||::::::|||   ||||||:::|||:::   :::|||::
 52 GlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAs 68

249 ATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAAC. 297
    ::::::::::::::|||||||||||||||||||||||:::::::   |||
 68 nAlaLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnA 85

298 ..GGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT 345
        :::   :::::::::      ||||||   |||||||||   |||
 85 rgArgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHis 101

346 AACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGA 395
        |||   :::         :::|||   ::::::   :::   ||
102 ValLysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAs 118

396 T......CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATA 439
    |     :::   :::      |||:::   |||   :::|||:::||||:
118 pValSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAspTyrG 135

440 AAGCAGGGACTAAAGGCCATCCTTATGGCGGC.............GAT 474
    ::   :::   :::   |||   :::|||:::              |||
135 luProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuGluAsp 150

475 TATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGA 524
    |||:::|||   ||||||:::||||||||||||:::::   |||:::
151 TyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaProIleAl 167

525 AATGACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAATAATT 571
        ||||||         |||   :::   |||   |||:::|||:::
167 aProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysAsnArgP 184

572 ACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGAT 621
    ::   :::   ||||||:::|||||||||   |||:::
184 heSerGluPheValArgValGlyAlaGlyThrGlnPhe............ 196

622 GAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTA 671
        |||   |||:::|||   :::      :::::::::   |||||
197 ......GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArgAlaTy 211

672 TTCTTGGCTCGTTGGTGGCAATACCTTT.............GCACAAA 706
    |   :::   :::::::|||:::   :::                :::::::|
211 rArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValThrSerA 228

707 ATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACAT 756
    ||   :::      |||   :::::::::|||:::::::::      |||
228 snLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLysHisHis 244

757 AGCCCA.....................TATGGTTTTTTACCAACAGG 782
    ||||||                        :::   |||   :::
245 SerProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuThrAsnTy 261

783 AGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAA 832
     :::   :::|||||||||||||||||:::|||   ||||||   |||:
261 rAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrAspLysGlnG 278

833 AGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATA 882
```

```
        ::::::::|||::::::   |||:::   :::          |||
278 luLysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGlyTyr... 293

883 GGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG...TTCTATGA 929
        ::::::|||::::::|||          :::|||   :::|||
294 ...GlnLysAsnSerTrpGln.....GluTrpAsnIleTyrLy 305

930 TGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAA 979
     :::   ||||||                      |||   :::|||:
305 sLysGluPheAlaAsp.....................GluIleLysGlnA 315

980 ATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAATCAAT 1029
    ::                    ||||||   ||||||   |||   |||
315 rg.....................AspAsnAlaGlyThrIleLysGlyAsn 324

1030 GCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGT 1079
     ::::::|||:::   ::::::                  |||   |||
325 GlyGluHisHisTrpLysThr.................ThrGlyThr.. 334

1080 TCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATC 1129
           |||       ::::::::::|||
335 ..........AsnSerHisIleGlySerThr..................... 341

1130 ATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAA 1179
     |||:::      :::|||:::   |||         ||||||||:::
342 ..AlaValArgLeuAlaAsnAsnGluArgAsp...AlaAsnAsnGlyGln 356

1180 AATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAA 1229
     |||::::::|||      |||::::::|||   |||:::|||   :::||
357 AsnValThrPhe......GluAsnAsnGlyThrLeuValLeuAspGlnAs 371

1230 CATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCT 1279
     ||||||||||||||||||||||||:::|||:::||||||:::|||||||:
371 nIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrThrValL 388

1280 CGCCTGAAAATAACGAA...ACTTGGCAAGGCGCGGGCGTTCATATCAGT 1326
     ::      ||||||:::   ||||||   ||||||||||:::   ::::::
388 ysGlyIleAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspValAla 404

1327 GAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCT 1376
     :::   :::   |||:::|||:::|||::::::      :::|||||||||
405 AspGlyLysLysValValTrpGlnValLysAsnProAsnGlyAspArgLe 421

1377 GTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACC 1426
     |:::|||||||||||||||||||||||||||:::::::::::   |||   ||||
421 uAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyValAsnG 438

1427 AAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCA 1476
     |||||:::::::::|||||||||||||||||||||||:::|||||||||
438 lnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnGlnAla 454

1477 GACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGG 1526
     |||         |||   ||||||||||||:::::|||:::|||||||||
455 AspAlaAspLysLysValGlnAlaPheSerGlnValGlyIleValSerGl 471

1527 CAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAAC 1576
     ||||||||||:::   ||||||:::::|||||:::||||||||||:::|
471 yArgGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAspAsnL 488

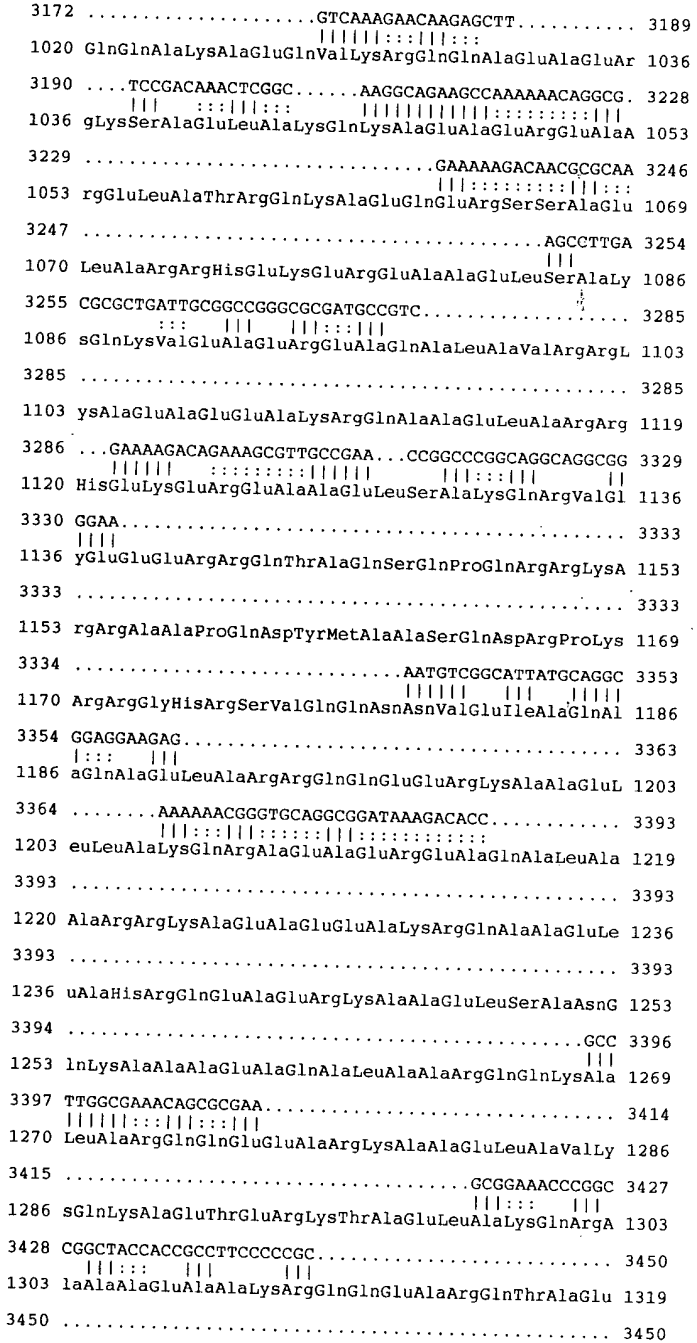
1577 TCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTT 1626
     |||||||||||||||||||||||||||||||   ||||||:::::|||
488 euTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeu 504

1627 TCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCA 1676
     :::|||:::::::|||:::|||:::||||||||||||   ||||||||||
505 ThrPheGluHisIleArgAsnValAspGluGlyAlaArgIleValAsnHi 521

1677 CAATCAAGACAAAGAATCCACCGTTACCATTACAGGC.............. 1713
     ||||          |||||||:::|||:::||||||
521 sAsnThrGlyHisAlaSerThrIleThrLeuThrGlyLysSerLeuIleT 538

1713 ..................................................... 1713
```




3172 .....GTCAAAGAACAAGAGCTT..... 3189
1020 GlnGlnAlaLysAlaGluGlnValLysArgGlnGlnAlaGluAlaGluAr 1036
3190 ....TCCGACAAACTCGGC......AAGGCAGAAGCCAAAAAACAGGCG. 3228
1036 gLysSerAlaGluLeuAlaLysGlnLysAlaGluAlaGluArgGluAlaA 1053
3229 .....GAAAAAGACAACGCGCAA 3246
1053 rgGluLeuAlaThrArgGlnLysAlaGluGlnGluArgSerSerAlaGlu 1069
3247 .....AGCCTTGA 3254
1070 LeuAlaArgArgHisGluLysGluArgGluAlaAlaGluLeuSerAlaLy 1086
3255 CGCGCTGATTGCGGCCGGGCGCGATGCCGTC..... 3285
1086 sGlnLysValGluAlaGluArgGluAlaGlnAlaLeuAlaValArgArgL 1103
3285 3285
1103 ysAlaGluAlaGluGluAlaLysArgGlnAlaAlaGluLeuAlaArgArg 1119
3286 ...GAAAAGACAGAAAGCGTTGCCGAA...CCGGCCCGGCAGGCAGGCGG 3329
1120 HisGluLysGluArgGluAlaAlaGluLeuSerAlaLysGlnArgValGl 1136
3330 GGAA..... 3333
1136 yGluGluGluArgArgGlnThrAlaGlnSerGlnProGlnArgArgLysA 1153
3333 3333
1153 rgArgAlaAlaProGlnAspTyrMetAlaAlaSerGlnAspArgProLys 1169
3334 .....AATGTCGGCATTATGCAGGC 3353
1170 ArgArgGlyHisArgSerValGlnGlnAsnAsnValGluIleAlaGlnAl 1186
3354 GGAGGAAGAG..... 3363
1186 aGlnAlaGluLeuAlaArgArgGlnGlnGluGluArgLysAlaAlaGluL 1203
3364 .....AAAAAACGGGTGCAGGCGGATAAAGACACC..... 3393
1203 euLeuAlaLysGlnArgAlaGluAlaGluArgGluAlaGlnAlaLeuAla 1219
3393 3393
1220 AlaArgArgLysAlaGluAlaGluGluAlaLysArgGlnAlaAlaGluLe 1236
3393 3393
1236 uAlaHisArgGlnGluAlaGluArgLysAlaAlaGluLeuSerAlaAsnG 1253
3394 .....GCC 3396
1253 lnLysAlaAlaAlaGluAlaGlnAlaLeuAlaAlaArgGlnGlnLysAla 1269
3397 TTGGCGAAACAGCGCGAA..... 3414
1270 LeuAlaArgGlnGlnGluGluAlaArgLysAlaAlaGluLeuAlaValLy 1286
3415 .....GCGGAAACCCGGC 3427
1286 sGlnLysAlaGluThrGluArgLysThrAlaGluLeuAlaLysGlnArgA 1303
3428 CGGCTACCACCGCCTTCCCCCGC..... 3450
1303 laAlaAlaGluAlaAlaLysArgGlnGlnGluAlaArgGlnThrAlaGlu 1319
3450 3450
1320 LeuAlaArgArgGlnGluAlaGluArgGlnAlaAlaGluLeuSerAlaLy 1336
3450 3450
1336 sGlnLysAlaGluThrAspArgGluAlaAlaGluSerAlaLysArgLysA 1353
3451 .....GCCCGCCGC 3459
1353 laGluGluGluGluHisArgGlnAlaAlaGlnSerGlnProGlnArgArg 1369
3460 GCCCGCCGGGATTTGCCGCAA..... 3480
1370 LysArgArgAlaAlaProGlnAspTyrMetAlaAlaSerGlnAsnArgPr 1386
3480 3480
1386 oLysArgArgGlyArgArgSerThrLeuProAlaProProSerProSerP 1403
3480 3480
1403 heAspSerSerAlaTyrAlaAlaProArgAlaLeuHisAsnProAspTrp 1419
3480 3480
1420 TyrGluAsnAspTyrGluGluIleProLeuAspAlaLeuGluAspGluAs 1436
3480 3480
1436 nValSerGluSerValAspThrSerAspLysGlnProGlnAspAsnThrG 1453
3480 3480
1453 luLeuHisGluLysTyrGluAsnAspTyrGluGluIleProLeuAspAla 1469
3480 3480
1470 LeuGluAspGluAspValSerGluSerValAspThrSerAspLysGlnPr 1486
3481 .....CTGCAAC 3487
1486 oGlnAspAsnThrGluLeuHisGluLysValGluThrValSerLeuGlnP 1503
3488 CC.....CAACCGCAG 3498
1503 roArgAlaAlaGlnProArgAlaGlnAlaAlaThrGlnLeuGlnAlaGln 1519
3499 CCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAG 3548
1520 AlaAlaAlaGlnAlaAspAlaValSerThrAsnThrAsnSerAlaLeuSe 1536
3549 TGAATTTTCCGCCACGCTCAACAGCGTTTTCGCC...GTACAGGACGAAT 3595
1536 rAspAlaMetAlaSerThrGlnSerIleLeuLeuAspThrGlyAlaSerL 1553
3596 TAGACCGCGTATTTGCCGAAGACCGCCGC.....AACGCC 3630
1553 euThrArgHisIleAlaGlnLysSerArgAlaAspAlaGluLysAsnSer 1569
3631 GTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAGATTT 3680
1570 ValTrpMetSerAsnThrGlyTyrGlyArgAspTyrAlaSerAlaGlnTy 1586
3681 CCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAA 3730
1586 rArgArgPheSerSerLysArgThrGlnThrGlnIleGlyIleAspArgS 1603
3731 ACCTCGGCAGCGGG...CGCGTCGGCATCCTGTTTTCGCACAACCGGACC 3777
1603 erLeuSerGluAsnMetGlnIleGlyGlyValLeuThrTyrSerAspSer 1619
3778 GAAAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGG 3827

```
1620 GlnHisThrPheAspGlnAlaSerGlyLysAsnThr...PheValGlnAl 1635

3828 CGCCGTTTTCGGGCAATACGGCATC...GACAGGTTCTACATCGGCATCA 3874
     :   :::::::|||:::|||   :::   |||   :::|||::::::   :
1635 aAsnLeuTyrGlyLysTyrTyrLeuAsnAspAlaTrpTyrValAlaGlyA 1652

3875 GCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGGC 3924
     ::   |||||||||          :::   |||:::            :::
1652 spIleGlyAlaGlySerLeuArgSerArgLeuGlnThrGlnGlnLysAla 1668

3925 AAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGC 3974
     :::::::::|||      ::::::   |||:::
1669 AsnPheAsnArgAlaSerIleGlnThrGlyLeuThrLeuGlyAsnThrLe 1685

3975 CGGTTTCGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATT 4024
          ::::::   |||   |||   |||      |||
1685 uLysIleAsnGlnPheGluIleValProSerAlaGlyIleArgTyrSerA 1702

4025 TCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCC 4074
       :::::::::|||||||||:::      |||   ::::::   :::
1702 rgLeuSerSerAlaAspTyrLysLeuGlyAsnAspSerValLysValSer 1718

4075 GGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATT 4124
     :::::::::   :::         ||||||:::         |||   ||
1719 SerMetSerValLysThrLeuThrAlaGlyLeuAspPheAlaTyrArgPh 1735

4125 CAAACCGGCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTAT. 4173
     ||||   :::   ::::::::::::   |||   ||||||   :::|||
1735 eLys...ValGlyAsnLeuThrValLysProLeuLeuSerAlaAlaTyrP 1751

4174 ..ACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACCGCCGTA 4221
       ::::::   ::::::|||   |||:::::      |||:::   |||
1751 heAlaAsnTyrGlyLysGlyGlyValAsnValGlyGlyAsnSerPheVal 1767

4222 TTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAACGC 4271
          |||      :::     ||||||      |||:::
1768 TyrLysAlaAspAsnGlnGlnGlnTyrSerAla......GlyAlaAlaLe 1782

4272 CGAAATCAAAGGTTTCACGCTGTCCCTCCACGCTGCCGCCGCCAAAGGCC 4321
           ::::::   ||||||:::::::::::::::   :::|||||||
1782 uLeuTyrArgAsnValThrLeuAsnValAsnGlySerIleThrLysGlyL 1799

4322 CGCAACTGGAAGCGCAACACAGCGCGGGCATCAAATTAGGCTACCGCTGG 4371
       |||||||||   |||   |||:::   |||||||:::      |||:::
1799 ysGlnLeuGluLysGlnLysSerGlyGlnIleLysIleGlnIleArgPhe 1815
```

```
seq_name: sp_bacteria:Q93T34

seq_documentation_block:
ID   Q93T34      PRELIMINARY;      PRT;  1764 AA.
AC   Q93T34;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE TYPE 2.
GN   IGA1.
OS   Haemophilus aegyptius.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=725;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=F3031;
RA   McGillivary G., Actis L.A.;
RT   "Iga protease from H. aegyptius F3031.";
RL   Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF369907; AAK56925.1; -.
KW   Protease.
SQ   SEQUENCE   1764 AA;  193863 MW;  EC583CCDE81DBE6C CRC64;


alignment_scores:
            Quality: 1455.00            Length:   1937
                Ratio:  1.573                 Gaps:     56
 Percent Similarity:  47.754   Percent Identity: 24.058

alignment_block:
 US-09-303-518D-649 x Q93T34   ..

 Align seg 1/1  to: Q93T34  from: 1  to: 1764
```

```
  64 CGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCC 113
     :::|||:::      ::::::|||:::   ::::::::::::::   ||
   5 LysPheLysLeuAsnPheIleAlaLeuThrValAlaTyrAlaLeuThrPr 21

 114 CCAAGCCTGGGCGGGACACACTTATTTCGGCATCAACTACCAATACTATC 163
     |   :::   |||:::   :::          ::::::||||||   :::|
  21 oTyrThrGluAlaAlaLeuValArgAspAspValAspTyrGlnIlePheA 38

 164 GCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATATT 213
     ||||||||||||||||||||||||:::|||:::||||||||||   ::::::
  38 rgAspPheAlaGluAsnLysGlyArgPheSerValGlyAlaThrAsnVal 54

 214 GAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAGC 263
     ||||||   :::|||:::::::::   :::|||:::::::::   :::::
  55 GluValArgAspLysAsnAsnHisSerLeuGlyAsnAlaLeuProAsnGl 71

 264 C...CCGATGATTGATTTTTCTGTGGTGTCGCGTAAC...GGCGTGGCGG 307
     :   ||||||||||||||||||||||:::   |||      :::::::
  71 yIleProMetIleAspPheSerValValAspValAsnLysArgIleGlyT 88

 308 CATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTAT 357
     ::||||||      ||||||||||||||||||   |||         |||
  88 hrLeuValAspProGlnTyrIleValSerValLysHisAlaHisGlnTyr 104

 358 AACAACGTTGATTTT.................GGTGCGGAAGGAAGAAA 389
         |||   ||||||                |||   :::      ::
 105 MetAsn...AspPheTyrPheGlyHisTyrAsnGlyHisArgAspValSe 120

 390 TCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATA 439
     :   |||:::::::::      |||:::::|||   :::||||||   :
 120 rAsnAspGluAsnLys......TyrSerValValThrGlnAsnAsnValA 135

 440 AAGCAGGGACTAAAGGC......CATCCTTATGGCGGCGATTATCATATG 483
     ::::::::   |||         :::              ||||||:::|||
 135 snSerSerGluLysTrpAspValAsnLysArgLeuAspAspTyrAsnMet 151

 484 CCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAG 533
     |||||||||:::||||||||||||:::::::   |||:::      |||
 152 ProArgLeuAsnLysPheValThrGluValAlaProThrThrProThrLe 168

 534 TTATATGGATGGG...CGGAAATATATCGATCAAAATAATTACCCTGACC 580
           |||      :::   |||   |||:::::::::||||||:::
 168 uAlaGlyAspAspLeuGluThrTyrLysAspLysGluLysTyrProSerP 185

 581 GTGTTCGTATTGGGGCAGGCAGGCAA......TATTGGCGATCTGATGAA 624
       ||||||:::|||||||||||||||      |||   ::::::::::::
 185 heValArgValGlyAlaGlyArgGlnLeuValTyrGluLysGlySerArg 201

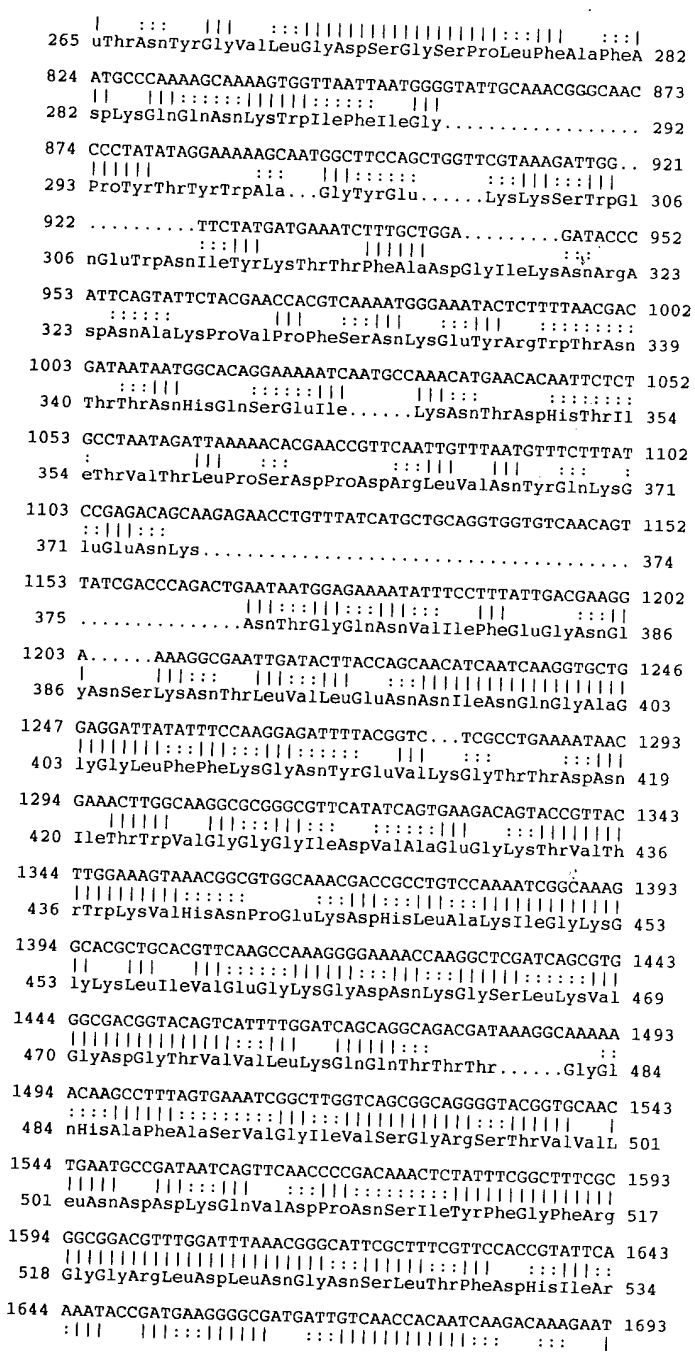
 625 GATGAGCCCAATAACCGC......GAAAGTTCATATCATATTGCAAGTGC 668
               :::|||:::      |||::         ::::::   ||
 202 HisValGluGlyAsnGluHisGlyGluAspLeuLysAspLeuSerValAl 218

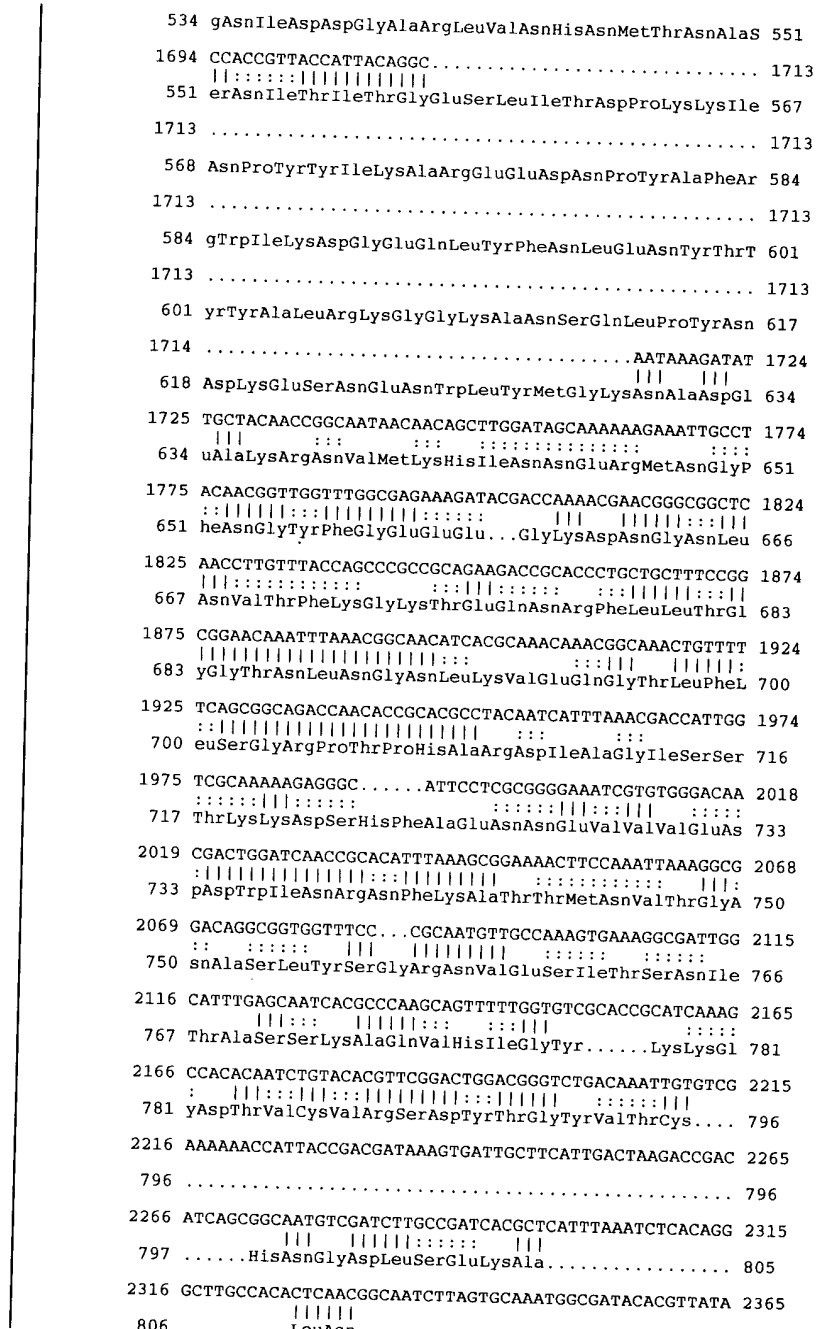
 669 GTATTCTTGGCTCGTTGGTGGCAATACCTTTGCA.............C 703
     ||||::::::   :::||||||:::   :::
 218 aTyrAsnTyrAlaIleGlyGlyThrProTyrGluGlyIleAsnIleAspP 235

 704 AAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAA 753
       :::   |||      |||   :::::::::|||:::::::::
 235 roSerGlnSerLysLysGlyLeuIleGlyPheGlyAspSerArgLysAsp 251

 754 CAT.........................AGCCCATATGGTTTTTT 773
     |||                        :::|||        ||
 252 HisValIleAspThrLysIleLeuLeuSerGlnAlaPro.....Le 265

 774 ACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATG 823
```


265 uThrAsnTyrGlyValLeuGlyAspSerGlySerProLeuPheAlaPheA 282
824 ATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAAC 873
282 spLysGlnGlnAsnLysTrpIlePheIleGly 292
874 CCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG 921
293 ProTyrThrTyrTrpAla...GlyTyrGlu.....LysLysSerTrpGl 306
922 .....TTCTATGATGAAATCTTTGCTGGA.....GATACCC 952
306 nGluTrpAsnIleTyrLysThrThrPheAlaAspGlyIleLysAsnArgA 323
953 ATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGAC 1002
323 spAsnAlaLysProValProPheSerAsnLysGluTyrArgTrpThrAsn 339
1003 GATAATAATGGCACAGGAAAAATCAATGCCAAACATGAACACAATTCTCT 1052
340 ThrThrAsnHisGlnSerGluIle.....LysAsnThrAspHisThrIl 354
1053 GCCTAATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTAT 1102
354 eThrValThrLeuProSerAspProAspArgLeuValAsnTyrGlnLysG 371
1103 CCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGT 1152
371 luGluAsnLys..... 374
1153 TATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGG 1202
375 .....AsnThrGlyGlnAsnValIlePheGluGlyAsnGl 386
1203 A.....AAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTG 1246
386 yAsnSerLysAsnThrLeuValLeuGluAsnAsnIleAsnGlnGlyAlaG 403
1247 GAGGATTATATTTCCAAGGAGATTTTACGGTC...TCGCCTGAAAATAAC 1293
403 lyGlyLeuPhePheLysGlyAsnTyrGluValLysGlyThrThrAspAsn 419
1294 GAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTAC 1343
420 IleThrTrpValGlyGlyGlyIleAspValAlaGluGlyLysThrValTh 436
1344 TTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAG 1393
436 rTrpLysValHisAsnProGluLysAspHisLeuAlaLysIleGlyLysG 453
1394 GCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTG 1443
453 lyLysLeuIleValGluGlyLysGlyAspAsnLysGlySerLeuLysVal 469
1444 GGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAA 1493
470 GlyAspGlyThrValValLeuLysGlnGlnThrThrThr.....GlyGl 484
1494 ACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAAC 1543
484 nHisAlaPheAlaSerValGlyIleValSerGlyArgSerThrValValL 501
1544 TGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGC 1593
501 euAsnAspAspLysGlnValAspProAsnSerIleTyrPheGlyPheArg 517
1594 GGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCA 1643
518 GlyGlyArgLeuAspLeuAsnGlyAsnSerLeuThrPheAspHisIleAr 534
1644 AAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAAT 1693


534 gAsnIleAspAspGlyAlaArgLeuValAsnHisAsnMetThrAsnAlaS 551
1694 CCACCGTTACCATTACAGGC..... 1713
551 erAsnIleThrIleThrGlyGluSerLeuIleThrAspProLysLysIle 567
1713 1713
568 AsnProTyrTyrIleLysAlaArgGluGluAspAsnProTyrAlaPheAr 584
1713 1713
584 gTrpIleLysAspGlyGluGlnLeuTyrPheAsnLeuGluAsnTyrThrT 601
1713 1713
601 yrTyrAlaLeuArgLysGlyGlyLysAlaAsnSerGlnLeuProTyrAsn 617
1714 .....AATAAAGATAT 1724
618 AspLysGluSerAsnGluAsnTrpLeuTyrMetGlyLysAsnAlaAspGl 634
1725 TGCTACAACCGGCAATAACAACAGCTTGGATAGCAAAAAAGAAATTGCCT 1774
634 uAlaLysArgAsnValMetLysHisIleAsnAsnGluArgMetAsnGlyP 651
1775 ACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGCTC 1824
651 heAsnGlyTyrPheGlyGluGluGlu...GlyLysAspAsnGlyAsnLeu 666
1825 AACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGG 1874
667 AsnValThrPheLysGlyLysThrGluGlnAsnArgPheLeuLeuThrGl 683
1875 CGGAACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTT 1924
683 yGlyThrAsnLeuAsnGlyAsnLeuLysValGluGlnGlyThrLeuPheL 700
1925 TCAGCGGCAGACCAACACCGCACGCCTACAATCATTTAAACGACCATTGG 1974
700 euSerGlyArgProThrProHisAlaArgAspIleAlaGlyIleSerSer 716
1975 TCGCAAAAAGAGGGC.....ATTCCTCGCGGGGAAATCGTGTGGGACAA 2018
717 ThrLysLysAspSerHisPheAlaGluAsnAsnGluValValValGluAs 733
2019 CGACTGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCG 2068
733 pAspTrpIleAsnArgAsnPheLysAlaThrThrMetAsnValThrGlyA 750
2069 GACAGGCGGTGGTTTCC...CGCAATGTTGCCAAAGTGAAAGGCGATTGG 2115
750 snAlaSerLeuTyrSerGlyArgAsnValGluSerIleThrSerAsnIle 766
2116 CATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAG 2165
767 ThrAlaSerSerLysAlaGlnValHisIleGlyTyr.....LysLysGl 781
2166 CCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCG 2215
781 yAspThrValCysValArgSerAspTyrThrGlyTyrValThrCys.... 796
2216 AAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGAC 2265
796 796
2266 ATCAGCGGCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGG 2315
797 .....HisAsnGlyAspLeuSerGluLysAla..... 805
2316 GCTTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATA 2365
806 .....LeuAsn..... 807



```
1312 laGluProAlaGlnGluAsnSerProThrValAsnAsnValGluGluAla 1328

3484 CAACCCCAACCGCAGCCCCAA..................................... 3504
1329 GlnProGlnThrGlnProThrIleValAlaAlaLysGluIleThrAlaPr 1345

3504 ............................................................ 3504

1345 oAsnSerAlaGlnLysGluThrAlaGlnSerValAlaAsnProLysThrA 1362

3504 ............................................................ 3504

1362 laGluGlnProValThrValSerThrGluAsnProValValGluAsnPro 1378

3504 ............................................................ 3504

1379 GluAsnThrThrGlnProAlaValAsnSerGluAlaValGlnProGluTh 1395

3504 ............................................................ 3504

1395 rAlaThrThrGlyThrValSerGlnProLysGluAlaSerThrAspGluT 1412

3504 ............................................................ 3504

1412 hrThrValAlaSerThrAspGluThrThrGlyThrSerAlaGluGluThr 1428

3504 ............................................................ 3504

1429 ThrValAlaAspAsnSerGluAlaSerLysProLysArgArgSerArgAr 1445

3504 ............................................................ 3504

1445 gAspValSerSerThrProHisAsnValGluProAlaValThrGlyGlyG 1462

3505 .....................CCGCAGCGCGACCTGATCAGCCGT 3528
1462 lyArgAspArgSerAlaValValValProLeuArgAspLeuThrSerThr 1478

3529 TATGCCAATAGCGGTTTGAGTGAATTTTCCGCC.............AC 3563
1479 AsnThrAsnAlaValLeuSerAspAlaMetAlaLysAlaGlnPheValAl 1495

3564 GCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTA...TTTG 3610
1495 aLeuAsnValGlyLysAlaValSerGlnHisIleSerGlnLeuGluMetA 1512

3611 CCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAA 3660
1512 snAsnGluGlyGlnTyrAsnValTrpValSerAsnThrSerMetAsnGlu 1528

3661 CACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCG 3710
1529 AsnTyrSerSerAspGlnTyrArgArgPheSerSerLysSerAlaGlnTh 1545

3711 CCAAATCGGTATGCAGAAAAACCTCGGCAGCGGG...CGCGTCGGCATCC 3757
1545 rGlnLeuGlyTrpAspGlnThrIleSerAsnAsnValGlnLeuGlyGlyI 1562

3758 TGTTTTCGCACAACCGGACCGAAAACACCTTCGACGACGGCATCGGCAAC 3807
1562 leLeuThrTyrValArgAsnSerAsnSerPheAspLysAlaSerSerLys 1578

3808 TCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGAC.. 3855
1579 AsnThr...LeuAlaGlnAlaAsnPheTyrSerLysTyrTyrAlaAspAs 1594

3856 .AGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCC 3904
1594 nHisTrpTyrLeuAlaValAspLeuGlyTyrGlyAsnPheGlnSerAsnL 1611

3905 TTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGC 3954

1611 euGlnThrAsnHisAsnAlaLysPheAspArgHisThrAlaGlnIleGly 1627

3955 ATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCA 4004
1628 LeuThrAlaGlyLysAlaPheAsnLeuGlyAsnValAlaValLysProTh 1644

4005 CATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAA 4054
1644 rValGlyValArgTyrSerTyrLeuSerAsnAlaAspPheAlaLeuAspG 1661

4055 ACGTCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGC 4104
1661 lnAspArgIleLysValAsnProIleSerValLysThrAlaPheAlaGln 1677

4105 ATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCC 4154
1678 ValAspLeuSerTyrThrTyrAsnLeuGlyGlu...PheAlaIleThrPr 1693

4155 TTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACAC 4204
1693 oIleLeuSerAlaArgTyr...AspAlaAsnGlnGlyAsnGlyLysIleA 1709

4205 GCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCG 4254
1709 snValSerGlyTyrAlaPheAlaTyrAsnValGluAsnGlnGlnGlnTyr 1725

4255 GAATGGGGCGTAAACGCCGAAATCAAAGGTTTCACGCTGTCCCTCCACGC 4304
1726 AsnAlaGlyLeuLysLeuLysTyrHisAsnValLysLeuSerLeuMetGl 1742

4305 TGCCGCCGCCAAAGGCCCGCAACTGGAAGCGCAACACAGCGCGGGCATCA 4354
1742 yGlyLeuThrLysAlaLysGlnAlaGluLysGlnLysThrAlaGluValL 1759

4355 AATTAGGCTAC 4365
1759 ysLeuSerPhe 1762
```

seq_name: sp_bacteria:Q57309

seq_documentation_block:

```
ID   Q57309       PRELIMINARY;      PRT;   996 AA.
AC   Q57309;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE PRECURSOR (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z3524;
RX   MEDLINE=98010345; PubMed=9350862;
RA   Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
RA   Achtman M.;
RT   "Clonal descent and microevolution of Neisseria meningitidis during 30
RT   years of epidemic spread.";
RL   Mol. Microbiol. 25:1047-1064(1997).
RN   [2]
RP   SEQUENCE OF 37-532 FROM N.A.
RC   STRAIN=ETH2;
RX   MEDLINE=95302961; PubMed=7783620;
RA   Lomholt H., Poulsen K., Mogens K.;
RT   "Comparative characterization of the iga gene encoding IgA1 protease
RT   in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
RT   influenzae.";
RL   Mol. Microbiol. 15:495-506(1995).
DR   EMBL; AF012207; AAC45790.2; -.
DR   EMBL; X82469; CAA57852.1; -.
DR   EMBL; X82468; CAA57851.1; -.
DR   MEROPS; S06.001; -.
```

```
DR   InterPro; IPR002195; Dihydrooratase.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
DR   PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     996    996
SQ   SEQUENCE   996 AA;  109717 MW;  AB39F7B98A41F986 CRC64;


alignment_scores:
            Quality: 1390.00              Length:   1188
              Ratio:  2.078                Gaps:     38
 Percent Similarity:  56.313   Percent Identity: 31.566

alignment_block:
 US-09-303-518D-649 x Q57309   ..

 Align seg 1/1  to: Q57309  from: 1  to: 996

      94 CTGTCGTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGG 143
         |||:::::::::::::   |||   :::   |||:::   :::
       1 LeuAlaTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAs 17

     144 CATCAACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTG 193
         ::::::||||||   :::||||||||||||||||||||||||||||||||
      17 pValAspTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPheP 34

     194 CAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC 243
         |||||||||   |||:::::::|||   ||||||:::|||:::   :::
      34 heValGlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIle 50

     244 GGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCG 293
         |||:::::::::::::::::::|||||||||||||||||||||||:::::
      51 GlyAsnAlaLeuSerAsnValProMetIleAspPheSerValAlaAspVa 67

     294 TAAC...GGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGG 340
         |||      :::   :::::::::      |||||||   |||||||||
      67 lAsnArgArgThrLeuThrValIleAspProGlnTyrAlaValSerValL 84

     341 CACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAAT 390
         |||      |||   :::          :::|||   ::::::   :::
      84 ysHisValLysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHis 100

     391 CCCGAT......CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAA 434
         |||      :::   :::      |||:::   |||   ::::|||::
     101 LeuAspValSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAs 117

     435 TTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGC.............. 471
         :|||:::   :::   :::   |||   :::|||:::
     117 pTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuG 133

     472 ..GATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCT 519
          ||||||:::|||   ||||||:::||||||||||||:::::   |||
     133 luAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaPro 149

     520 GTTGAAATGACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAA 566
         :::      ||||||          |||   :::   |||   |||:::||
     150 IleAlaProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysAs 166

     567 TAATTACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGAT 616
         |::::::   :::   ||||||:::||||||||   |||:::
     166 nArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe....... 180

     617 CTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGT 666
                    |||   |||:::|||   :::      :::::::::
     181 ..........GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArg 193

     667 GCGTATTCTTGGCTCGTTGGTGGCAATACCTTT.............GC 701
         ||||||   :::   ::::::|||:::   :::              ::
     194 AlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValTh 210

     702 ACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTA 751
         ::::|||   :::      |||   :::::::::|||:::::::::
     210 rSerAsnLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLysH 227

     752 AACATAGCCCA.....................TATGGTTTTTTACCA 777
           |||||||||                     :::   |||
     227 isHisSerProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuThr 243

     778 ACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGC 827
         :::   :::   :::|||||||||||||||||||:::|||   ||||||
     244 AsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrAspLy 260

     828 CCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCT 877
         |||::::::::::|||::::::   |||:::   :::             |
     260 sGlnGluLysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGlyT 277

     878 ATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG...TTC 924
         ||      ::::::|||::::::|||            :::|||   :::
     277 yr......GlnLysAsnSerTrpGln............GluTrpAsnIle 287

     925 TATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACG 974
         |||   :::   ||||||                              |||   ::
     288 TyrLysLysGluPheAlaAsp.....................GluIleLy 297

     975 TCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAA 1024
         :|||:::                  ||||||   ||||||   |||
     297 sGlnArg.....................AspAsnAlaGlyThrIleLysG 307

    1025 TCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGA 1074
               ::::::|||:::   ::::::                  |||
     307 lyTyrGlyGluHisHisTrpLysThr..................ThrGly 317

    1075 ACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGT 1124
         |||            |||      :::::::::||||||   :::   ::
     318 Thr............AsnSerHisIleGlySerThrAlaValArgLeuAl 330

    1125 TTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATG 1174
         :   :::      |||:::                        |||||||
     330 aGlyAsnGluArgGlyAla.......................AsnAsnG 339

    1175 GAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACC 1224
         ||:::|||::::::|||      |||   :::|||   |||:::|||
     339 lyGlnAsnValThrPhe......GluAspAsnGlyThrLeuValLeuAsp 353

    1225 AGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTAC 1274
         :::||||||||||||||||||||||||||:::|||:::|||||||:::||
     354 GlnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrTh 370

    1275 GGTCTCGCCTGAAAATAACGAA...ACTTGGCAAGGCGCGGGCGTTCATA 1321
         ||||:::      ||||||:::   ||||||   ||||||||||:::   :
     370 rValLysGlyAlaAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspV 387

    1322 TCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGAC 1371
         ::::::::   :::   |||:::|||:::|||::::::      :::|||
     387 alAlaAspGlyLysLysValValTrpGlnValLysAsnProAsnGlyAsp 403

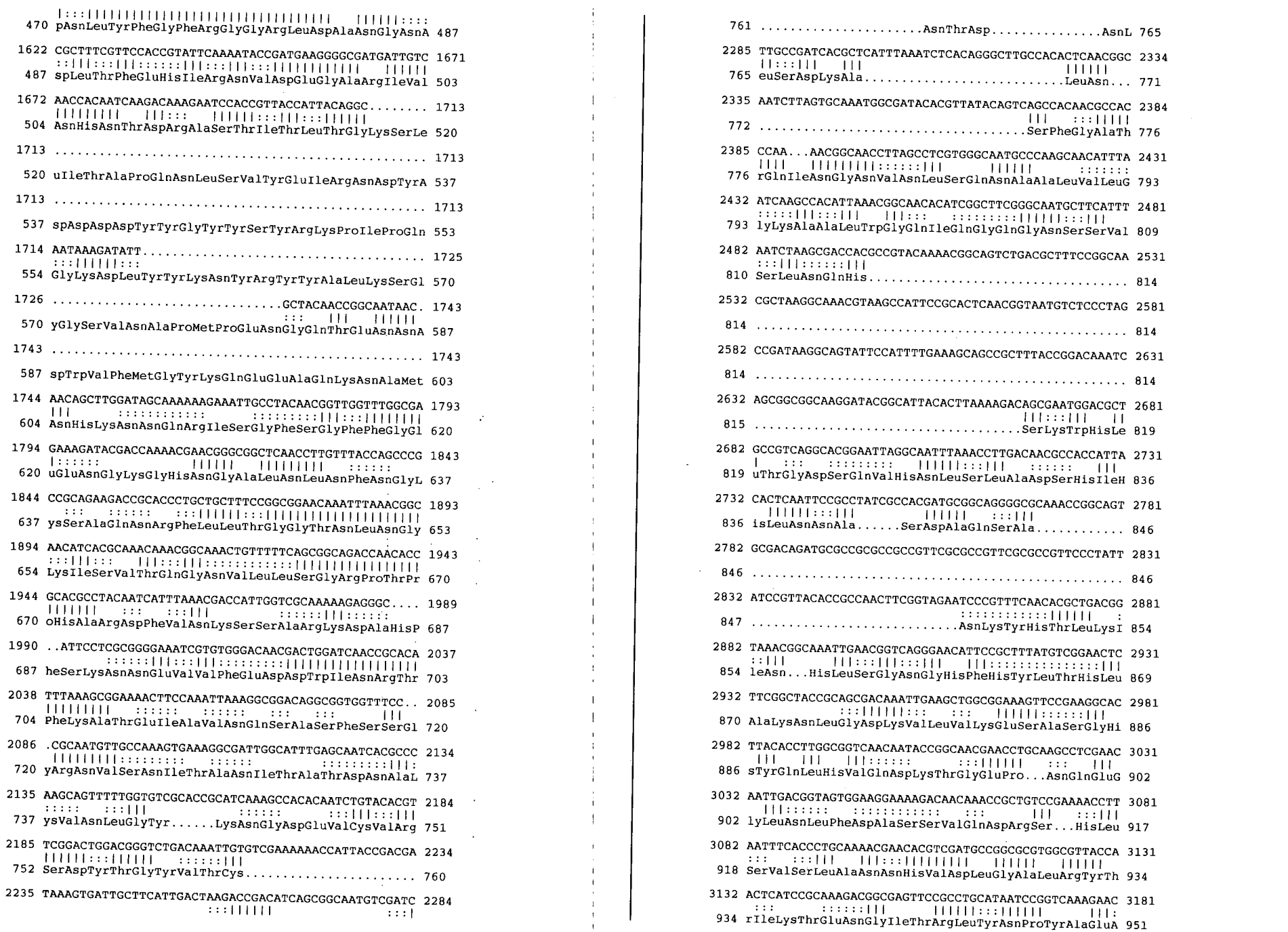
    1372 CGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGA 1421
         ||||||:::||||||||||||||||||||||:::::::::::   |||
     404 ArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyVa 420

    1422 AAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGC 1471
         ||||||||:::::::::||||||||||||||||||||||||||:::||||
     420 lAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnG 437

    1472 AGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTC 1521
         ||||||||      |||   ||||||||||||:::::::|||:::|||
     437 lnAlaAspAlaAspLysLysValGlnAlaPheSerGlnValGlyIleVal 453

    1522 AGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGA 1571
         ||||||||||||||::   ||||||:::::::||||||:::||||||||
     454 SerGlyArgGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAs 470

    1572 CAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATT 1621
```



```
   3182 AAGAGCTTTCCGACAAACTCGGC.....................AAGGCA 3210
        :::::        |||   :::                          :::
    951 snArgArgArgValLysProAlaProSerProAlaThrAsnThrAlaSer 967

   3211 GAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCT 3260
        :::|||:::|||   ::::::   |||:::||||||
    968 GlnAlaGlnLysAlaThrGlnThrAspGlyAlaGln.............. 979

   3261 GATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAAC 3310
         ||||||      ::::::   |||               ||||||   |
    980 .IleAlaLysProGlnAsnIleVal..............ValAlaProP 991

   3311 CGGCCCGGCAGGCA 3324
        ||:::   ||||||
    991 roSerProGlnAla 995

seq_name: sp_bacteria:O30575

seq_documentation_block:
ID   O30575      PRELIMINARY;      PRT;   997 AA.
AC   O30575;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z4099;
RX   MEDLINE=98010345; PubMed=9350862;
RA   Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
RA   Achtman M.;
RT   "Clonal descent and microevolution of Neisseria meningitidis during 30
RT   years of epidemic spread.";
RL   Mol. Microbiol. 25:1047-1064(1997).
DR   EMBL; AF012210; AAC45793.2; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR002195; Dihydrooratase.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
DR   PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     997    997
SQ   SEQUENCE   997 AA;  110152 MW;  4DE40DE594E5E28E CRC64;


alignment_scores:
            Quality: 1383.50              Length:   1187
              Ratio:  2.068                Gaps:     37
 Percent Similarity:  56.361   Percent Identity: 31.171

alignment_block:
 US-09-303-518D-649 x O30575   ..

 Align seg 1/1  to: O30575  from: 1  to: 997

      94 CTGTCGTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGG 143
         |||:::::::::::::   |||   :::   |||:::   :::
       1 LeuAlaTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAs 17

     144 CATCAACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTG 193
         :::::||||||   :::||||||||||||||||||||||||||||||||
      17 pValAspTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPheP 34

     194 CAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC 243
           |||||||||   |||:::::::|||   ||||||:::|||:::   :::
      34 heValGlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIle 50

     244 GGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCG 293
         |||:::::::::::::::::::|||||||||||||||||||||||:::::
      51 GlyAsnAlaLeuSerAsnValProMetIleAspPheSerValAlaAspVa 67

     294 TAAC...GGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGG 340
          |||      :::   :::::::::      ||||||   |||||||||
      67 lAsnArgArgThrLeuThrValIleAspProGlnTyrAlaValSerValL 84

     341 CACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAAT 390
         |||      |||   :::         :::|||   ::::::   :::
      84 ysHisValLysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHis 100

     391 CCCGAT......CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAA 434
         |||      :::   :::      |||:::   |||   :::|||::
     101 LeuAspValSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAs 117

     435 TTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGC............. 471
         :|||:::   :::   :::   |||   :::|||:::
     117 pTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuG 133

     472 ..GATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCT 519
           ||||||:::|||   ||||||:::||||||||||||:::::::   |||
     133 luAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaPro 149

     520 GTTGAAATGACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAA 566
         :::      ||||||         |||   :::   |||   |||:::||
     150 IleAlaProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysAs 166

     567 TAATTACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGAT 616
         |::::::   :::   ||||||:::|||||||||   |||:::
     166 nArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe....... 180

     617 CTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGT 666
                   |||   |||:::|||   :::       :::::::::
     181 ..........GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArg 193

     667 GCGTATTCTTGGCTCGTTGGTGGCAATACCTTT..............GC 701
         ||||||   :::   ::::::|||:::   :::              ::
     194 AlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValTh 210

     702 ACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTA 751
         ::::|||   :::      |||   :::::::::|||:::::::::::
     210 rSerAsnLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLysH 227

     752 AACATAGCCCA.....................TATGGTTTTTTACCA 777
           |||||||||                          :::   |||
     227 isHisSerProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuThr 243

     778 ACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGC 827
         :::   :::   :::||||||||||||||||||:::|||   ||||||
     244 AsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrAspLy 260

     828 CCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCT 877
          |||:::::::::|||::::::   |||:::   :::          |
     260 sGlnGluLysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGlyT 277

     878 ATATAGGAAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG...TTC 924
         ||      ::::::|||::::::|||            :::|||   :::
     277 yr......GlnLysAsnSerTrpGln............GluTrpAsnIle 287

     925 TATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACG 974
         |||   :::   ||||||                          |||   ::
     288 TyrLysLysGluPheAlaAsp.....................GluIleLy 297

     975 TCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAA 1024
         :|||:::                  ||||||   ||||||   |||
     297 sGlnHis.....................AspAsnAlaGlyThrIleLysG 307

    1025 TCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGA 1074
                ::::|||::   |||              :::   :::
     307 lyTyrArgGluHisHisTrpAsn..............IleThrSerGly 318
```



```
 816 ..............................SerLysTrpHisLeuTh 821

2685 GTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACAC 2734
      :::   :::::::::   ||||||:::|||   ::::::   |||   |
 821 rGlyAspSerGlnValHisAsnLeuSerLeuAlaAspSerHisIleHisL 838

2735 TCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCG 2784
     |||||:::|||         ||||||   :::|||
 838 euAsnAsnAla......SerAspAlaGlnSerAla..............  847

2785 ACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATC 2834

 847 ................................................. 847

2835 CGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAA 2884
                                 ::::::::::::||||||   :::|
 848 ........................AsnLysTyrHisThrLeuLysIleA 856

2885 ACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTC 2934
     ||      |||:::|||:::|||   |||:::::::::::::::|||
 856 sn...HisLeuSerGlyAsnGlyHisPheHisTyrLeuThrHisLeuAla 871

2935 GGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTA 2984
               :::|||||||:::   :::   ||||||:::::::|||   ||
 872 GluAsnLeuGlyAspLysValLeuValLysGluSerAlaSerGlyHisTy 888

2985 CACCTTGGCGGTCAACAATACCGGCAACGAACCTGCAAGCCTCGAACAAT 3034
     |   |||   |||::::::      :::||||||   :::   |||:::|
 888 rGlnLeuHisValGlnAspLysThrGlyGluPro...AsnGlnGluArgL 904

3035 TGACGGTAGTGGAAGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAAT 3084
     ||::::::   ::::::::::::   :::      |||   :::|||:::
 904 euAsnLeuPheAspAlaSerSerValArgAspArgSer...ArgLeuSer 919

3085 TTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACT 3134
        :::|||   |||:::||||||||||   ||||||   ||||||   ::
 920 ValSerLeuAlaAsnAsnHisValAspLeuGlyAlaLeuArgTyrThrIl 936

3135 CATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAG 3184
     :      ::::::|||      ||||||:::||||||      ||||::::
 936 eLysThrGluAsnGlyIleThrArgLeuTyrAsnProTyrAlaGluAsnA 953

3185 AGCTTTCCGACAAA.....................CTCGGCAAGGCAGAA 3213
     ::   :::   |||                     ::::::
 953 rgArgGlnValLysProThrHisSerProAlaThrAsnThrAlaSerGln 969

3214 GCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGAT 3263
     |||:::|||   ::::::   |||:::||||||              ||
 970 AlaGlnLysAlaThrGlnThrAspGlyAlaGln..............Il 981

3264 TGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGG 3313
     ||||      ::::::   |||              ||||||   |||:
 981 eAlaLysProGlnAsnIleVal..............ValAlaProProS 993

3314 CCCGGCAGGCA 3324
     ::   ||||||
 993 erProGlnAla 996

seq_name: sp_bacteria:Q57035

seq_documentation_block:
ID   Q57035      PRELIMINARY;      PRT;   992 AA.
AC   Q57035;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE PRECURSOR (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=B40;
RX   MEDLINE=98010345; PubMed=9350862;
RA   Morelli G., Malorny B., Mueller K., Seiler A., Wang J.-F.,
RA   del Valle J., Achtman M.;
RT   "Clonal descent and microevolution of Neisseria meningitidis during 30
RT   years of epidemic spread.";
RL   Mol. Microbiol. 25:1047-1064(1997).
RN   [2]
RP   SEQUENCE OF 37-532 FROM N.A.
RC   STRAIN=HF48;
RX   MEDLINE=95302961; PubMed=7783620;
RA   Lomholt H., Poulsen K., Mogens K.;
RT   "Comparative characterization of the iga gene encoding IgA1 protease
RT   in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
RT   influenzae.";
RL   Mol. Microbiol. 15:495-506(1995).
DR   EMBL; AF012211; AAC45794.2; -.
DR   EMBL; X82480; CAA57863.1; -.
DR   EMBL; X82475; CAA57858.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR002195; Dihydrooratase.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
DR   PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     992    992
SQ   SEQUENCE   992 AA;  109228 MW;  3677DDE4CE6D9F69 CRC64;


alignment_scores:
            Quality: 1380.00              Length:   1183
              Ratio:  2.072                Gaps:     35
 Percent Similarity:  56.298   Percent Identity: 31.107

alignment_block:
 US-09-303-518D-649 x Q57035   ..

 Align seq 1/1  to: Q57035  from: 1  to: 992

      94 CTGTCGTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGG 143
         |||:::::::::::::   |||   :::   ||:::   :::
       1 LeuAlaTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAs 17

     144 CATCAACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTG 193
          ::::::||||||   :::|||||||||||||||||||||||||||||||
      17 pValAspTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPheP 34

     194 CAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC 243
           |||||||||   |||:::::::|||   ||||||:::|||:::   :::
      34 heValGlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIle 50

     244 GGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCG 293
         |||:::::::::::::::::::|||||||||||||||||||||||::::::
      51 GlyAsnAlaLeuSerAsnValProMetIleAspPheSerValAlaAspVa 67

     294 TAAC...GGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGG 340
          |||      :::   :::::::::      ||||||   |||||||||
      67 lAsnArgArgThrLeuThrValIleAspProGlnTyrAlaValSerValL 84

     341 CACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAAT 390
           |||      |||   :::          :::|||   ::::::   :::
      84 ysHisValLysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHis 100

     391 CCCGAT.....CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAA 434
           |||      :::   :::      |||:::   |||   :::|||::
     101 LeuAspValSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAs 117

     435 TTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGC............. 471
         :|||:::   :::   :::   |||   :::|||:::
     117 pTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuG 133
```

```
 472 ..GATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCT 519
     ||||||:::|||  ||||||:::||||||||||||:::::  |||
 133 luAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaPro 149

 520 GTTGAAATGACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAA 566
     :::    ||||||        |||  :::  |||  |||:::||
 150 IleAlaProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysAs 166

 567 TAATTACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGAT 616
     |::::::  :::  ||||||:::|||||||||  |||:::
 166 nArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe..... 180

 617 CTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGT 666
               |||  |||:::|||  :::      :::::::::
 181 ..........GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArg 193

 667 GCGTATTCTTGGCTCGTTGGTGGCAATACCTTT.............GC 701
     ||||||  :::  :::::::|||:::  :::             ::
 194 AlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValTh 210

 702 ACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTA 751
     ::::|||  :::      |||  :::::::::|||:::::::::
 210 rSerAsnLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLysH 227

 752 AACATAGCCCA.....................TATGGTTTTTTACCA 777
       ||||||||                    :::  |||
 227 isHisSerProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuThr 243

 778 ACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGC 827
     :::  :::  ::::|||||||||||||||||||:::|||  ||||||
 244 AsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrAspLy 260

 828 CCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCT 877
      |||:::::::::|||::::::  |||:::  :::           |
 260 sGlnGluLysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGlyT 277

 878 ATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG...TTC 924
     ||      :::::::|||:::::|||           :::|||  :::
 277 yr......GlnLysAsnSerTrpGln...........GluTrpAsnIle 287

 925 TATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACG 974
     |||  :::  ||||||                    |||  ::
 288 TyrLysLysGluPheAlaAsp.....................GluIleLy 297

 975 TCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAA 1024
     :|||:::                 ||||||  ||||||  |||
 297 sGlnArg.....................AspAsnAlaGlyThrIleLysG 307

1025 TCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGA 1074
       ::::::|||:::  ::::::                  |||
 307 lyTyrGlyGluHisHisTrpLysThr.................ThrGly 317

1075 ACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGT 1124
     |||      |||  :::::::::||||||  :::  ::
 318 Thr...........AsnSerHisIleGlySerThrAlaValArgLeuAl 330

1125 TTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATG 1174
     :  :::      |||:::                 |||||||
 330 aGlyAsnGluArgGlyAla.....................AsnAsnG 339

1175 GAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACC 1224
     ||:::|||:::::|||      |||::::::|||  |||:::|||
 339 lyGlnAsnValThrPhe......GluAsnAsnGlyThrLeuValLeuAsp 353

1225 AGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTAC 1274
     :::||||||||||||||||||||||||:::|||:::||||||:::||
 354 GlnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrTh 370

1275 GGTCTCGCCTGAAAATAACGAA...ACTTGGCAAGGCGCGGGCGTTCATA 1321
     ||||:::      ||||||:::  ||||||  ||||||||||:::  :
 370 rValLysGlyIleAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspV 387

1322 TCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGAC 1371

     ::::::::  :::  |||:::|||:::|||::::::      :::|||
 387 alAlaAspGlyLysLysValValTrpGlnValLysAsnProAsnGlyAsp 403

1372 CGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGA 1421
     ||||||:::||||||||||||||||||||:::::::::::  |||
 404 ArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyVa 420

1422 AAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGC 1471
     |||||||||:::::::::||||||||||||||||||||||:::|||:
 420 lAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnL 437

1472 AGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTC 1521
     ::||||||::::::  |||  ||||||||||||:::::|||:::|||
 437 ysAlaAspSerAsnGlnLysValGlnAlaPheSerGlnValGlyIleVal 453

1522 AGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGA 1571
     ||||||||||||||::  ||||||:::::|||||||:::|||||||
 454 SerGlyArgGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAs 470

1572 CAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATT 1621
     |:::|||||||||||||||||||||||||||||||  ||||||::::
 470 pAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnA 487

1622 CGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTC 1671
     ::|||:::|||:::::|||:::|||:::||||||||||||  ||||||
 487 spLeuThrPheGluHisIleArgAsnValAspGluGlyAlaArgIleVal 503

1672 AACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGC........ 1713
     |||||||||            ||||||:::|||:::||||||
 504 AsnHisAsnThrGlyHisThrSerThrIleThrLeuThrGlyLysSerLe 520

1713 ..................................................... 1713
 520 uIleThrAsnProAsnSerLeuSerValHisSerIleGlnAsnAspTyrA 537

1713 ..................................................... 1713
 537 spGluAspAspTyrSerTyrTyrTyrArgProArgArgProIleProGln 553

1714 AATAAAGATATT..................................... 1725
     :::|||||||:::
 554 GlyLysAspLeuTyrTyrLysAsnTyrArgTyrTyrAlaLeuLysSerGl 570

1726 ............................GCTACAACCGGCAATAAC. 1743
                                ::::  |||  ||||||
 570 yGlySerValAsnAlaProMetProGluAsnGlyGlnThrGluAsnAsnA 587

1743 ..................................................... 1743
 587 spTrpIleLeuMetGlySerThrGlnGluGluAlaLysLysAsnAlaMet 603

1744 AACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGA 1793
     |||      ::::::::::::      :::::::::|||:::||||||||
 604 AsnHisLysAsnAsnGlnArgIleSerGlyPheSerGlyPhePheGlyGl 620

1794 GAAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCG 1843
     |::::::              ||||||  |||||||||  ::::::
 620 uGluAsnGlyLysGlyHisAsnGlyAlaLeuAsnLeuAsnPheAsnGlyL 637

1844 CCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGC 1893
       :::  ::::::  :::||||||:::||||||||||||||||||||
 637 ysSerAlaGlnAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAsnGly 653

1894 AACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAACACC 1943
     :::|||:::  |||:::|||:::::::::::::|||||||||||||||
 654 LysIleSerValThrGlnGlyAsnValLeuLeuSerGlyArgProThrPr 670

1944 GCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGC.... 1989
     |||||||  :::  :::|||         ::::::|||::::::
 670 oHisAlaArgAspPheValAsnLysSerSerAlaArgLysAspAlaHisP 687

1990 ..ATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACA 2037
           ::::::|||:::|||::::::::::|||||||||||||||||||
```

```
 687 heSerLysAsnAsnGluValValPheGluAspAspTrpIleAsnArgThr 703

2038 TTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCC.. 2085
     |||||||||   ::::::   ::::::   :::   :::       |||
 704 PheLysAlaThrGluIleAlaValAsnGlnSerAlaSerPheSerSerGl 720

2086 .CGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCC 2134
      |||||||||:::   :::   ::::::          :::::::::|||:
 720 yArgAsnValSerAspIleThrAlaAsnIleThrAlaThrAspAsnAlaL 737

2135 AAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGT 2184
     :::::   :::|||           ::::::      :::|||:::|||
 737 ysValAsnLeuGlyTyr......LysAsnGlyAspGluValCysValArg 751

2185 TCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACCGACGA 2234
     ||||||:::||||||   :::::::|||         :::::::::   ||
 752 SerAspTyrThrGlyTyrValThrCysAsnThrGlyAsnLeuSer...As 767

2235 TAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTCGATC 2284
     ||||::::::   |||:::      |||:::|||:::||||||||||:::|
 767 pLysAlaLeuAsnSerPheGlyAlaThrGlnIleAsnGlyAsnValAsnL 784

2285 TTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGC 2334
     ||::::::
 784 euSerGln..................................... 786

2335 AATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCAC 2384

 786 ..................................................... 786

2385 CCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATC 2434
                                  ||||||      |||:::::::
 787 .........................AsnAlaAlaLeuThrLeuGlyL 794

2435 AAGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTTAAT 2484
     ::|||:::|||   |||:::   :::::::::||||||:::      :::
 794 ysAlaAlaLeuTrpGlyGlnIleGlnGlyGlnGlyAsnSerArgValSer 810

2485 CTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAACGC 2534
     |||::::::|||
 811 LeuAsnGlnHis..................................... 814

2535 TAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTAGCCG 2584

 814 ..................................................... 814

2585 ATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGC 2634

 814 ..................................................... 814

2635 GGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCC 2684
                                      |||:::|||   |||
 815 ...............................SerLysTrpHisLeuTh 820

2685 GTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACAC 2734
      :::   :::::::::   ||||||:::|||   ::::::   |||   |
 820 rGlyAspSerGlnValHisAsnLeuSerLeuAlaAspSerHisIleHisL 837

2735 TCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCG 2784
     |||||:::|||         ||||||   :::|||
 837 euAsnAsnAla......SerAspAlaGlnSerAla.............. 846

2785 ACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATC 2834

 846 ..................................................... 846

2835 CGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAA 2884
                                 ::::::::::::||||||   :::|
 847 ........................AsnLysTyrHisThrLeuLysIleA 855

2885 ACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTC 2934
     ||      |||:::|||:::|||   |||:::::::::::::::|||
 855 sn...HisLeuSerGlyAsnGlyHisPheHisTyrLeuThrHisLeuAla 870

2935 GGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTA 2984
               :::|||||||:::   :::   ||||||:::::::|||   ||
 871 LysAsnLeuGlyAspLysValLeuValLysGluSerAlaSerGlyHisTy 887

2985 CACCTTGGCGGTCAACAATACCGGCAACGAACCTGCAAGCCTCGAACAAT 3034
     |   |||   |||::::::      :::|||||   :::   |||   |
 887 rGlnLeuHisValGlnAspLysThrGlyGluPro...AsnGlnGluGlyL 903

3035 TGACGGTAGTGGAAGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAAT 3084
     ||:::::   :::::::::::   :::      |||   :::|||:::
 903 euAsnLeuPheAspAlaSerSerValArgAspArgSer...HisLeuSer 918

3085 TTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACT 3134
        :::|||   |||:::|||||||||   ||||||   ||||||   ::
 919 ValSerLeuAlaAsnAsnHisValAspLeuGlyAlaLeuArgTyrThrIl 935

3135 CATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAG 3184
     :     :::::::|||      ||||||:::||||||      |||::::
 935 eLysThrGluAsnGlyIleThrArgLeuTyrAsnProTyrAlaGluAsnA 952

3185 AGCTTTCCGACAAACTCGGCAAGGCAGAAGCC.........AAAAAACAG 3225
     ::        |||   :::   :::   |||              :::|||
 952 rgArgArgValLysProAlaProSerProAlaThrAsnThrAlaSerGln 968

3226 GCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCG 3275
     |||:::   |||:::||||||                ||||||      ::
 969 AlaGlnThrAspSerAlaGln..............IleAlaLysProGl 980

3276 CGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCA 3324
     ::::   |||               ||||||   |||:::   ||||||
 980 nAsnIleVal..............ValAlaProProSerProGlnAla 991
```

```
seq_name: sp_bacteria:Q9S6X4

seq_documentation_block:
ID   Q9S6X4      PRELIMINARY;      PRT;   992 AA.
AC   Q9S6X4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z3906;
RX   MEDLINE=98010345; PubMed=9350862;
RA   Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
RA   Achtman M.;
RT   "Clonal descent and microevolution of Neisseria meningitidis during 30
RT   years of epidemic spread.";
RL   Mol. Microbiol. 25:1047-1064(1997).
DR   EMBL; AF012204; AAC45787.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR002195; Dihydrooratase.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
DR   PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     992    992
SQ   SEQUENCE   992 AA;  109239 MW;  9ED0A41019917CE3 CRC64;


alignment_scores:
            Quality: 1376.00              Length:   1183
              Ratio:  2.069                Gaps:     36
 Percent Similarity:  56.213   Percent Identity: 31.276
```

```
alignment_block:
 US-09-303-518D-649 x Q9S6X4   ..

 Align seg 1/1  to: Q9S6X4  from: 1  to: 992

      94 CTGTCGTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGG 143
         |||:::::::::::   |||   :::   |||:::   :::
       1 LeuAlaTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAs 17

     144 CATCAACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTG 193
         ::::::||||||   :::|||||||||||||||||||||||||||||||
      17 pValAspTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPheP 34

     194 CAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC 243
           |||||||||   |||:::::::|||   ||||||:::|||:::   :::
      34 heValGlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIle 50

     244 GGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCG 293
         |||:::::::::::::::::::|||||||||||||||||||||:::|::
      51 GlyAsnAlaLeuSerAsnValProMetIleAspPheSerValAlaAspVa 67

     294 TAAC...GGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGG 340
          |||      :::   :::::::::      ||||||   |||||||||
      67 lAsnArgArgThrLeuThrValIleAspProGlnTyrAlaValSerValL 84

     341 CACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAAT 390
          |||      |||   :::          :::|||   ::::::   :::
      84 ysHisValLysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHis 100

     391 CCCGAT......CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAA 434
          |||      :::   :::      |||:::   |||   :::||::
     101 LeuAspValSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAs 117

     435 TTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGC..... 471
         :|||:::   :::   :::   |||   :::|||:::
     117 pTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuG 133

     472 ..GATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCT 519
           ||||||:::|||   ||||||:::||||||||||||:::::   |||
     133 luAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaPro 149

     520 GTTGAAATGACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAA 566
         :::      ||||||          |||   :::   |||   |||:::||
     150 IleAlaProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysAs 166

     567 TAATTACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGAT 616
         |::::::   :::   ||||||:::|||||||||   |||:::
     166 nArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe..... 180

     617 CTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGT 666
                 |||   |||:::|||   :::      :::::::::
     181 ..........GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArg 193

     667 GCGTATTCTTGGCTCGTTGGTGGCAATACCTTT.............GC 701
         ||||||   :::   ::::::|||:::   :::            ::
     194 AlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValTh 210

     702 ACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTA 751
         ::::|||   :::      |||   :::::::::|||:::::::::
     210 rSerAsnLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLysH 227

     752 AACATAGCCCA.....................TATGGTTTTTTACCA 777
           |||||||||                     :::   |||
     227 isHisSerProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuThr 243

     778 ACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGC 827
         :::   :::   :::|||||||||||||||||||:::|||   ||||||
     244 AsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrAspLy 260

     828 CCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCT 877
         |||:::::::::|||::::::   |||:::   :::          |
     260 sGlnGluLysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGlyT 277

     878 ATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG...TTC 924
         ||      :::::::|||:::::::|||             :::|||   :::
     277 yr......GlnLysAsnSerTrpGln............GluTrpAsnIle 287

     925 TATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACG 974
         |||   :::   ||||||                              |||   ::
     288 TyrLysLysGluPheAlaAsp.....................GluIleLy 297

     975 TCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAA 1024
         :|||:::                    ||||||   ||||||   |||
     297 sGlnArg.....................AspAsnAlaGlyThrIleLysG 307

    1025 TCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGA 1074
           ::::::|||:::   ::::::                     |||
     307 lyTyrGlyGluHisHisTrpLysThr..................ThrGly 317

    1075 ACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGT 1124
         |||            |||      :::::::::||||||   :::   ::
     318 Thr............AsnSerHisIleGlySerThrAlaValArgLeuAl 330

    1125 TTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATG 1174
         :   :::      |||:::                        |||||||
     330 aGlyAsnGluArgGlyAla.......................AsnAsnG 339

    1175 GAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACC 1224
         ||:::|||::::::|||      |||::::::|||   |||:::|||
     339 lyGlnAsnValThrPhe......GluAsnAsnGlyThrLeuValLeuAsp 353

    1225 AGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTAC 1274
         :::||||||||||||||||||||||||||||:::|||:::||||||:::||
     354 GlnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrTh 370

    1275 GGTCTCGCCTGAAAATAACGAA...ACTTGGCAAGGCGCGGGCGTTCATA 1321
         ||||:::      ||||||      ||||||   |||||||||:::   :
     370 rValLysGlyAlaAsnAsnGlyIleThrTrpLeuGlyAlaGlyIleAspV 387

    1322 TCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGAC 1371
         ::::::::   :::   |||:::|||:::|||::::::      :::|||
     387 alAlaAspGlyLysLysValValTrpGlnValLysAsnProAsnGlyAsp 403

    1372 CGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGA 1421
         ||||||:::||||||||||||||||||||||:::::::::::   |||
     404 ArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyVa 420

    1422 AAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGC 1471
          ||||||||||:::::::::||||||||||||||||||||||||:::||||
     420 lAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnG 437

    1472 AGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTC 1521
         ||||||||          |||   ||||||||||:::::|||:::|||
     437 lnAlaAspAlaAspLysLysValGlnAlaPheSerGlnValGlyIleVal 453

    1522 AGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGA 1571
         ||||||||||||||::   ||||||:::   :::|||:::||||||::
     454 SerGlyArgGlyThrLeuValLeuAsnSerProAspGlnIleAsnProAs 470

    1572 CAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATT 1621
         ::::|||||||||||||||||||||||||||||||   ||||||:::
     470 nAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnA 487

    1622 CGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTC 1671
         ::|||:::|||:::::::|||:::|||:::||||||||||||   ||||||
     487 spLeuThrPheGluHisIleArgAsnValAspGluGlyAlaArgIleVal 503

    1672 AACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGC..... 1713
         |||||||||   |||:::   ||||||:::|||:::||||||
     504 AsnHisAsnThrAspArgAlaSerThrIleThrLeuThrGlyLysSerLe 520

    1713 ..................................................... 1713

     520 uIleThrAlaProGlnAsnLeuSerValTyrGluIleArgAsnAspTyrA 537

    1713 ..................................................... 1713
```



seq_name: sp_bacteria:Q9S6X3

seq_documentation_block:

```
ID   Q9S6X3      PRELIMINARY;      PRT;   993 AA.
AC   Q9S6X3;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z4400;
RX   MEDLINE=98010345; PubMed=9350862;
RA   Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
RA   Achtman M.;
RT   "Clonal descent and microevolution of Neisseria meningitidis during 30
RT   years of epidemic spread.";
RL   Mol. Microbiol. 25:1047-1064(1997).
DR   EMBL; AF012208; AAC45791.2; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR002195; Dihydrooratase.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
DR   PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     993    993
SQ   SEQUENCE   993 AA;  109441 MW;  109FAA2EF88AC3C6 CRC64;


alignment_scores:
            Quality: 1375.50              Length:   1186
              Ratio:  2.075                Gaps:     35
 Percent Similarity:  55.902   Percent Identity: 31.029

alignment_block:
 US-09-303-518D-649 x Q9S6X3   ..

 Align seg 1/1  to: Q9S6X3  from: 1  to: 993

      94 CTGTCGTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGG 143
         |||:::::::::::::   |||   :::   |||:::   :::
       1 LeuAlaTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAs 17

     144 CATCAACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTG 193
         ::::::||||||   :::||||||||||||||||||||||||||||||
      17 pValAspTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPheP 34

     194 CAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC 243
         |||||||||   |||:::::::|||   |||||||:::|||:::   :::
      34 heValGlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIle 50

     244 GGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCG 293
         |||:::::::::::::::::|||||||||||||||||||||::::::
      51 GlyAsnAlaLeuSerAsnValProMetIleAspPheSerValAlaAspVa 67

     294 TAAC...GGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGG 340
           |||      :::   :::::::::      ||||||   |||||||||
      67 lAsnArgArgThrLeuThrValIleAspProGlnTyrAlaValSerValL 84

     341 CACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAAT 390
         |||      |||   :::         :::|||   ::::::  :::
      84 ysHisValLysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHis 100

     391 CCCGAT......CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAA 434
         |||      :::   :::      |||:::   |||   :::|||::
     101 LeuAspValSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAs 117

     435 TTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGC..............  471
         :|||:::   :::   :::   |||   :::|||:::
     117 pTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuG 133

     472 ..GATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCT 519
           ||||||:::|||   ||||||:::|||||||||||||::::::   |||
     133 luAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaPro 149

     520 GTTGAAATGACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAA 566
         :::      ||||||          |||   :::   |||   |||:::||
     150 IleAlaProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysAs 166

     567 TAATTACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGAT 616
         |::::::   :::   ||||||:::|||||||||   |||:::
     166 nArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe....... 180

     617 CTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGT 666
                 |||   |||:::|||   :::       :::::::::
     181 ..........GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArg 193

     667 GCGTATTCTTGGCTCGTTGGTGGCAATACCTTT.............GC 701
         ||||||   :::   ::::::|||:::   :::             ::
     194 AlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValTh 210

     702 ACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTA 751
         ::::|||   :::      |||   ::::::::::|||:::::::::
     210 rSerAsnLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLysH 227

     752 AACATAGCCCA.....................TATGGTTTTTTACCA 777
           |||||||||                     :::   |||
     227 isHisSerProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuThr 243

     778 ACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGC 827
         :::   :::   :::|||||||||||||||||||:::|||   ||||||
     244 AsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrAspLy 260

     828 CCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCT 877
          |||:::::::::|||::::::   |||:::  :::             |
     260 sGlnGluLysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGlyT 277

     878 ATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG...TTC 924
         ||      ::::::|||::::::|||            :::|||   :::
     277 yr......GlnLysAsnSerTrpGln............GluTrpAsnIle 287

     925 TATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACG 974
         |||   :::   ||||||
     288 TyrLysLysGluPheAlaAspLysIleLys..................... 297

     975 TCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAA 1024
                                 :::   ||||||   ||||||   |||
     298 .....................GlnArgAspAsnAlaGlyThrIleLysG 307

    1025 TCAATGCCAAACATGAACACAAT.........TCTCTGCCTAATAGATTA 1065
           |||::::::|||:::   |||         :::      :::::::::
     307 lyAsnGlyGluHisHisTrpAsnIleThrPheGlyThrAsnSerLysIle 323

    1066 AAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAG 1115
         :::   :::|||:::|||
     324 GlySerThrAlaValArgLeu............................. 330

    1116 AGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGAC 1165
                           ||||||:::   ::::::
     331 ..................AlaGlyAsnGluLysAsp............A 337

    1166 TGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTG 1215
         |||||||||:::|||::::::|||      |||   :::|||   |||
     337 laAsnAsnGlyGlnAsnValThrPhe......GluAspAsnGlyThrLeu 351

    1216 ATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGG 1265
         :::|||   :::|||||||||||||||||||||||||||:::|||:::||
     352 ValLeuAspGlnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGl 368

    1266 AGATTTTACGGTCTCGCCTGAAAATAACGAA...ACTTGGCAAGGCGCGG 1312
         ||||:::|||||:::      ||||||:::   ||||||   |||||||
     368 yAspTyrThrValLysGlyIleAsnAsnAspIleThrTrpLeuGlyAlaG 385
```

```
1313 GCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTG 1362
     ||:::   :::::::::   :::   |||:::|||:::|||::::::
 385 lyIleAspValThrAspGlyLysLysValValTrpGlnValLysAsnPro 401

1363 GCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGC 1412
        :::|||||||||:::|||||||||||||||||||||||:::::::::::
 402 AsnGlyAspArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGl 418

1413 CAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTT 1462
     :   |||   |||||||||:::::::::|||||||||||||||||||||
 418 yThrGlyValAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleL 435

1463 TGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATC 1512
     ||:::|||||||||||||         |||   |||||||||||:::::::
 435 euAsnGlnGlnAlaAspAlaAspLysLysValGlnAlaPheSerGlnVal 451

1513 GGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTT 1562
     |||:::|||||||||||||||||:::   ||||||:::::|||||||::
 452 GlyIleValSerGlyArgGlyThrLeuValLeuAsnSerSerAsnGlnIl 468

1563 CAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAA 1612
     :|||||||||:::|||||||||||||||||||||||||||||||||   |
 468 eAsnProAspAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaA 485

1613 ACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCG 1662
     |||||:::::|||:::|||:::::|||:::|||:::|||||||||||||
 485 snGlyAsnAspLeuThrPheGluHisIleArgAsnValAspGluGlyAla 501

1663 ATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGG 1712
        ||||||||||||||   :::      ||||||:::|||:::|||||
 502 ArgIleValAsnHisAsnThrSerHisAlaSerThrIleThrLeuThrGl 518

1713 C..................................................... 1713
     |
 518 yLysSerLeuIleThrAsnProAsnSerLeuSerValHisSerIleGlnA 535

1713 ..................................................... 1713

 535 snAspTyrAspGluAspAspTyrSerTyrTyrTyrArgProArgArgPro 551

1714 .........AATAAAGATATT.............................. 1725
             :::|||||||:::
 552 IleProGlnGlyLysAspLeuTyrTyrLysAsnTyrArgTyrTyrAlaLe 568

1726 ..........................................GCTACAACCG 1735
                                                 ::::::|||
 568 uLysSerGlyGlySerValAsnAlaProMetProGluAsnGlyValThrG 585

1736 GCAATAAC.............................................. 1743
        ||||||
 585 luAsnAsnAspTrpValPheMetGlyTyrThrGlnGluGluAlaLysLys 601

1744 .........AACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTG 1784
             |||      ::::::::::::      ::::::::::|||::
 602 AsnAlaMetAsnHisLysAsnAsnGlnArgIleSerGlyPheSerGlyPh 618

1785 GTTTGGCGAGAAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTT 1834
     :|||||||||:::::          ||||||  |||||||||   :
 618 ePheGlyGluGluAsnGlyLysGlyHisAsnGlyAlaLeuAsnLeuAsnP 635

1835 ACCAGCCCGCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAAT 1884
     :::::      :::   ::::::   :::|||||||:::|||||||||||
 635 heAsnGlyLysSerAlaGlnAsnArgPheLeuLeuThrGlyGlyThrAsn 651

1885 TTAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAG 1934
     |||||||||:::|||:::   |||:::|||:::::::::::::|||||||
 652 LeuAsnGlyLysIleSerValThrGlnGlyAsnValLeuLeuSerGlyAr 668

1935 ACCAACACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAG 1984
     ||||||||||||||   :::   :::|||         :::::::|||:
 668 gProThrProHisAlaArgAspPheValAsnLysSerSerAlaArgLysA 685

1985 AGGGC......ATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATC 2028
```

```
     :::::            ::::::|||:::|||::::::::::|||||||||
 685 spAlaHisPheSerLysAsnAsnGluValValPheGluAspAspTrpIle 701

2029 AACCGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGT 2078
     ||||||||||||||||   ::::::  ::::::   :::   :::
 702 AsnArgThrPheLysAlaAlaGluIleAlaValAsnGlnSerAlaSerPh 718

2079 GGTTTCC...CGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCA 2125
        |||   |||||||||::::::::   ::::::         ::::
 718 eSerSerGlyArgAsnValSerAsnIleThrAlaAsnIleThrAlaThrA 735

2126 ATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATC 2175
     ::::::|||:::::   :::|||              ::::::      :::
 735 spAsnAlaLysValAsnLeuGlyTyr......LysAsnGlyAspGluVal 749

2176 TGTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCAT 2225
     |||:::||||||||||:::|||||   ::::::|||         :::::
 750 CysValArgSerAspTyrThrGlyTyrValThrCysAsnThrGlyAsnLe 766

2226 TACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCA 2275
     ::::   ||||||::::::   |||:::      |||:::|||:::||||
 766 uSer...AspLysAlaLeuAsnSerPheGlyAlaThrGlnIleAsnGlyA 782

2276 ATGTCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACA 2325
     |||||:::|||
 782 snValAsnLeu..................................... 785

2326 CTCAACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCA 2375

 785 ..................................................... 785

2376 CAACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAA 2425
                  |||   |||   :::|||||||
 786 .............AsnGlnAsnAlaAlaLeuVal..... 792

2426 CATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCT 2475
       ::::::::::|||:::|||   |||:::   :::::::::|||||||:::
 793 ..LeuGlyLysAlaAlaLeuTrpGlyGlnIleGlnGlyGlnGlyAsnSer 808

2476 TCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTC 2525
           :::|||::::::|||
 809 ArgValSerLeuAsnGlnHis.......................... 815

2526 CGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCT 2575

 815 ..................................................... 815

2576 CCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGA 2625

 815 ..................................................... 815

2626 CAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATG 2675
                                                    |||:::||
 816 ..........................................SerLysTr 818

2676 GACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCA 2725
     |   |||   :::   :::::::::   ||||||:::|||   ::::::
 818 pHisLeuThrGlyAspSerGlnValHisAsnLeuSerLeuAlaAspSerH 835

2726 CCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAAACC 2775
        |||   ||||||:::|||         ||||||   :::|||
 835 isIleHisLeuAsnAsnAla......SerAspAlaGlnSerAla..... 847

2776 GGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTC 2825

 847 ..................................................... 847

2826 CCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGC 2875
                                        :::::::::::::||||
 848 ................................AsnGlnTyrHisThrL 853

2876 TGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCG 2925
     ||   :::|||      |||:::|||:::|||   |||:::::::::::::
```

```
 853 euLysIleAsn...HisLeuSerGlyAsnGlyHisPheHisTyrLeuThr 868

2926 GAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGA 2975
     :::|||                :::|||||||:::   :::   ||||||:::::
 869 HisLeuAlaGluAsnLeuGlyAspLysValLeuValLysGluSerAlaSe 885

2976 AGGCACTTACACCTTGGCGGTCAACAATACCGGCAACGAACCTGCAAGCC 3025
     :|||   |||   |||   |||::::::      :::|||||||   :::
 885 rGlyHisTyrGlnLeuHisValGlnAspLysThrGlyGluPro...AsnG 901

3026 TCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAACCGCTGTCCGAA 3075
       |||   |||::::::   ::::::::::::   :::      |||
 901 lnGluGlyLeuAsnLeuPheAspAlaSerSerValGlnAspArgSer... 916

3076 AACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCG 3125
     :::|||:::   :::|||   |||:::||||||||||   ||||||   ||
 917 ArgLeuSerValSerLeuAlaAsnAsnHisValAspLeuGlyAlaLeuAr 933

3126 TTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCA 3175
     ||||   :::      :::::::|||      ||||||:::||||||
 933 gTyrThrIleLysThrGluAsnGlyIleThrArgLeuTyrAsnProTyrA 950

3176 AAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCC......... 3216
       |||::::::         |||         :::   |||
 950 laGluAsnArgArgArgValLysProValProSerProAlaThrAsnThr 966

3217 AAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGC 3266
       :::||||||:::   |||:::|||||                ||||||
 967 AlaSerGlnAlaGlnThrAspSerAlaGln.............IleAl 978

3267 GGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCC 3316
     |      ::::::   |||                ||||||   |||:::
 978 aLysProGlnAsnIleVal.............ValAlaProProSerP 990

3317 GGCAGGCA 3324
       ||||||
 990 roGlnAla 992
```

```
seq_name: sp_bacteria:Q9S6X2

seq_documentation_block:
ID   Q9S6X2      PRELIMINARY;      PRT;   997 AA.
AC   Q9S6X2;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z4081;
RX   MEDLINE=98010345; PubMed=9350862;
RA   Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
RA   Achtman M.;
RT   "Clonal descent and microevolution of Neisseria meningitidis during 30
RT   years of epidemic spread.";
RL   Mol. Microbiol. 25:1047-1064(1997).
DR   EMBL; AF012209; AAC45792.2; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR002195; Dihydrooratase.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
DR   PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     997    997
SQ   SEQUENCE   997 AA;  109812 MW;  06F2E361E7E202E0 CRC64;
```

```
alignment_scores:
           Quality: 1375.50              Length:   1190
             Ratio:  2.068                Gaps:     35
 Percent Similarity:  55.882   Percent Identity: 30.840

alignment_block:
 US-09-303-518D-649 x Q9S6X2   ..

 Align seg 1/1  to: Q9S6X2  from: 1  to: 997

  94 CTGTCGTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGG 143
     |||:::::::::::   |||   :::   |||:::   :::
   1 LeuAlaTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAs 17

 144 CATCAACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTG 193
       ::::::||||||   :::||||||||||||||||||||||||||||||
  17 pValAspTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPheP 34

 194 CAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC 243
       |||||||||   |||:::::::|||   ||||||:::|||:::   :::
  34 heValGlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIle 50

 244 GGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCG 293
     |||:::::::::::::::::::|||||||||||||||||||||::::::
  51 GlyAsnAlaLeuSerAsnValProMetIleAspPheSerValAlaAspVa 67

 294 TAAC...GGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGG 340
       |||      :::   :::::::::      ||||||   |||||||||
  67 lAsnArgArgThrLeuThrValIleAspProGlnTyrAlaValSerValL 84

 341 CACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAAT 390
       |||      |||   :::          :::|||   ::::::   :::
  84 ysHisValLysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHis 100

 391 CCCGAT......CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAA 434
       |||      :::   :::      |||:::   |||   :::|||::
 101 LeuAspValSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAs 117

 435 TTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGC.............. 471
     :|||:::   :::   :::   |||   :::|||:::
 117 pTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuG 133

 472 ..GATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCT 519
       ||||||:::|||   |||||:::||||||||||||:::::   |||
 133 luAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaPro 149

 520 GTTGAAATGACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAA 566
     :::      ||||||          |||   :::   |||   |||:::||
 150 IleAlaProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysAs 166

 567 TAATTACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGAT 616
     |::::::   :::   ||||||:::|||||||||   |||:::
 166 nArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe....... 180

 617 CTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGT 666
                 |||   |||:::|||   :::      :::::::::
 181 ..........GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArg 193

 667 GCGTATTCTTGGCTCGTTGGTGGCAATACCTTT..............GC 701
     ||||||   :::   ::::::|||:::   :::              ::
 194 AlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValTh 210

 702 ACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTA 751
     :::::|||   :::      |||   :::::::::|||::::::::::
 210 rSerAsnLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLysH 227

 752 AACATAGCCCA.....................TATGGTTTTTTACCA 777
       |||||||||                     :::   |||
 227 isHisSerProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuThr 243

 778 ACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGC 827
     :::   :::   :::|||||||||||||||||:::|||   ||||||
 244 AsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrAspLy 260
```



```
                  |||   |||   :::||||||
  786 .............AsnGlnAsnAlaAlaLeuVal..................... 792

 2426 CATTTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCT 2475
        :::::::::|||:::|||   |||:::   :::::::::||||||:::
  793 ..LeuGlyLysAlaAlaLeuTrpGlyGlnIleGlnGlyGlnGlyAsnSer 808

 2476 TCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTC 2525
            :::|||:::::::|||
  809 ArgValSerLeuAsnGlnHis..............................  815

 2526 CGGCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCT 2575

  815 .....................................................  815

 2576 CCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGA 2625

  815 .....................................................  815

 2626 CAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATG 2675
                                          |||:::||
  816 ..........................................SerLysTr 818

 2676 GACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCA 2725
        |||   :::   :::::::::   |||||||:::|||   ::::::
  818 pHisLeuThrGlyAspSerGlnValHisAsnLeuSerLeuAlaAspSerH 835

 2726 CCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAAACC 2775
        |||   ||||||:::|||          ||||||   :::|||
  835 isIleHisLeuAsnAsnAla......SerAspAlaGlnSerAla...... 847

 2776 GGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTC 2825

  847 .....................................................  847

 2826 CCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGC 2875
                                        :::::::::::||||
  848 ................................AsnGlnTyrHisThrL 853

 2876 TGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCG 2925
        ||   :::|||      |||:::|||:::|||   |||:::::::::::
  853 euLysIleAsn...HisLeuSerGlyAsnGlyHisPheHisTyrLeuThr 868

 2926 GAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGA 2975
        :::|||           :::|||||||:::   :::   ||||||:::::
  869 HisLeuAlaGluAsnLeuGlyAspLysValLeuValLysGluSerAlaSe 885

 2976 AGGCACTTACACCTTGGCGGTCAACAATACCGGCAACGAACCTGCAAGCC 3025
        :|||   |||   |||   |||:::::::      :::||||||   :::
  885 rGlyHisTyrGlnLeuHisValGlnAspLysThrGlyGluPro...AsnG 901

 3026 TCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAACCGCTGTCCGAA 3075
        |||   |||::::::   ::::::::::::   :::      |||
  901 lnGluGlyLeuAsnLeuPheAspAlaSerSerValGlnAspArgSer... 916

 3076 AACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCG 3125
        :::|||:::   :::|||   |||:::|||||||||   ||||||   ||
  917 ArgLeuSerValSerLeuAlaAsnAsnHisValAspLeuGlyAlaLeuAr 933

 3126 TTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCA 3175
        ||||   :::       :::::::|||       ||||||:::||||||
  933 gTyrThrIleLysThrGluAsnGlyIleThrArgLeuTyrAsnProTyrA 950

 3176 AAGAACAAGAGCTTTCCGACAAACTCGGC..................... 3204
        |||::::::          |||   :::
  950 laGluAsnArgArgArgValLysProAlaProSerProAlaThrAsnThr 966

 3205 AAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGA 3254
        ::::::|||:::|||   ::::::   |||:::||||||
  967 AlaSerGlnAlaGlnLysAlaThrGlnThrAspGlyAlaGln........ 980

 3255 CGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTG 3304
            ||||||      ::::::   |||                ||||
```

```
  981 .......IleAlaLysProGlnAsnIleVal..............ValA 990

 3305 CCGAACCGGCCCGGCAGGCA 3324
        ||   |||:::   ||||||
  990 laProProSerProGlnAla 996

seq_name: sp_bacteria:Q9S6X5

seq_documentation_block:
ID   Q9S6X5      PRELIMINARY;      PRT;   992 AA.
AC   Q9S6X5;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z2491;
RX   MEDLINE=98010345; PubMed=9350862;
RA   Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
RA   Achtman M.;
RT   "Clonal descent and microevolution of Neisseria meningitidis during 30
RT   years of epidemic spread.";
RL   Mol. Microbiol. 25:1047-1064(1997).
DR   EMBL; AF012203; AAC45786.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR002195; Dihydrooratase.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
DR   PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     992    992
SQ   SEQUENCE   992 AA;  109239 MW;  306FF451B7213D49 CRC64;


alignment_scores:
            Quality: 1375.00              Length:   1183
              Ratio:  2.074                 Gaps:     37
 Percent Similarity:  56.044   Percent Identity: 31.361

alignment_block:
US-09-303-518D-649 x Q9S6X5   ..

 Align seg 1/1  to: Q9S6X5  from: 1  to: 992

   94 CTGTCGTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGG 143
      |||:::::::::::   |||   :::   |||:::   :::
    1 LeuAlaTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAs 17

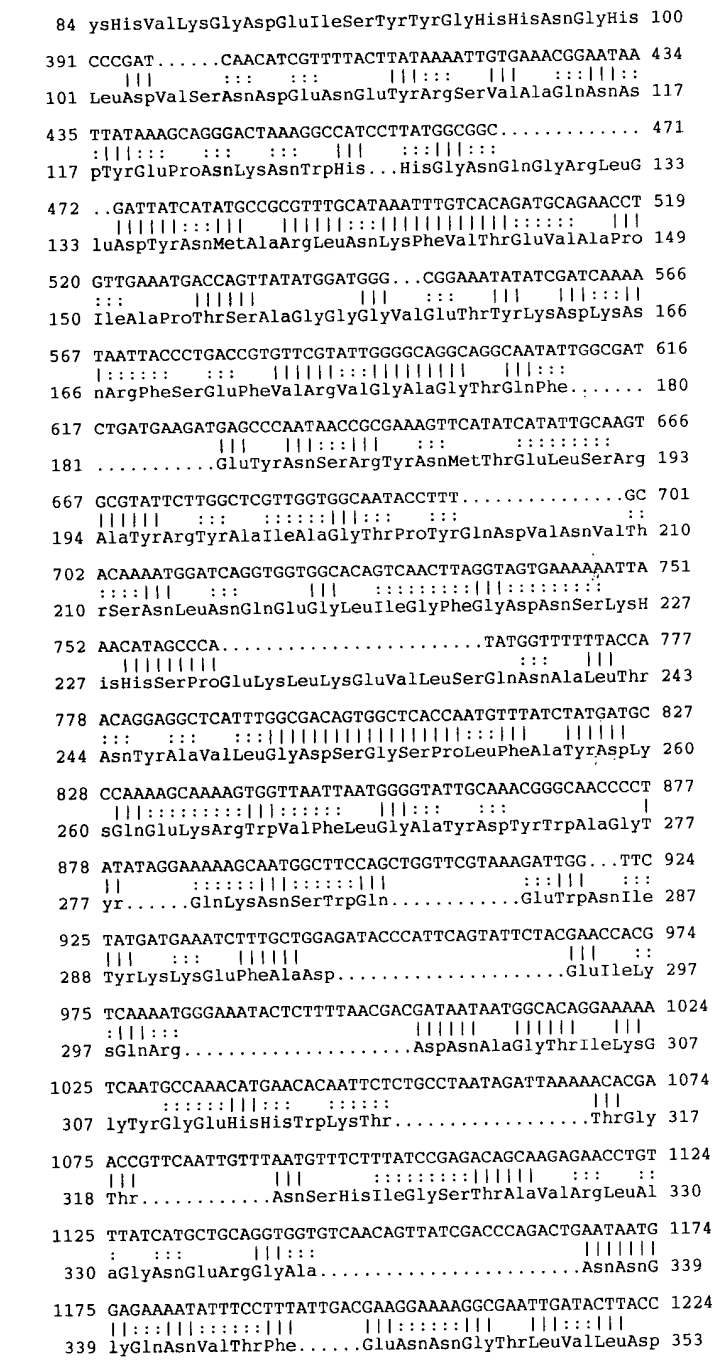
  144 CATCAACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTG 193
        ::::::||||||   :::|||||||||||||||||||||||||||||
   17 pValAspTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPheP 34

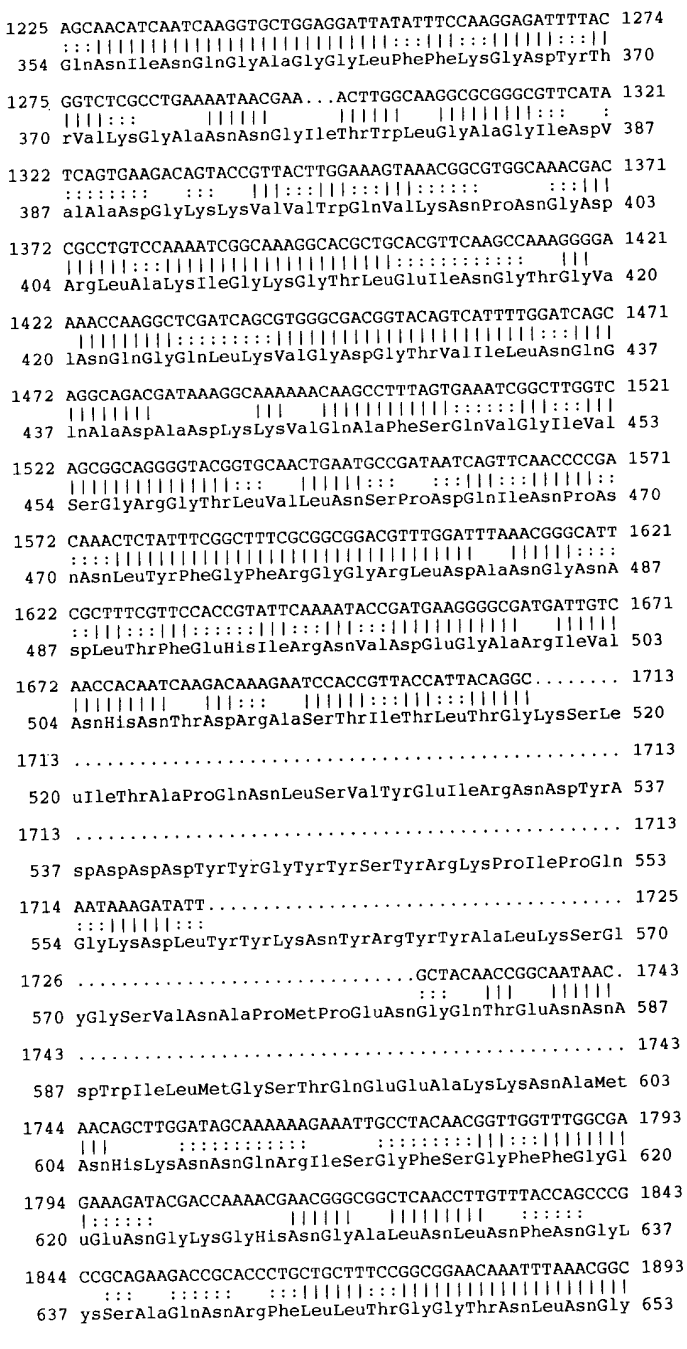
  194 CAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC 243
        |||||||||   |||:::::::|||   ||||||:::|||:::   :::
   34 heValGlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIle 50

  244 GGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCG 293
      |||:::::::::::::::::::|||||||||||||||||||||::::::
   51 GlyAsnAlaLeuSerAsnValProMetIleAspPheSerValAlaAspVa 67

  294 TAAC...GGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGG 340
        |||      :::   :::::::::      ||||||   |||||||||
   67 lAsnArgArgThrLeuThrValIleAspProGlnTyrAlaValSerValL 84

  341 CACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAAT 390
        |||      |||   :::         :::|||   ::::::   :::
```





```
1894 AACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAACACC 1943
      :::|||:::   |||:::|||:::::::::::::|||||||||||||||||
 654 LysIleSerValThrGlnGlyAsnValLeuLeuSerGlyArgProThrPr 670

1944 GCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGC.... 1989
      |||||||   :::   :::|||         :::||||||:::::
 670 oHisAlaArgAspPheValAsnLysSerSerAlaGlnLysAspAlaHisP 687

1990 ..ATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACA 2037
            ::::::|||:::|||:::::::::|||||||||||||||||||
 687 heSerLysAsnAsnGluValValPheGluAspAspTrpIleAsnArgThr 703

2038 TTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCC.. 2085
      |||||||||   ::::::   ::::::   :::   :::      |||
 704 PheLysAlaThrGluIleThrValAsnGlnSerAlaSerPheSerSerGl 720

2086 .CGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCC 2134
      |||||||||:::::::::   ::::::         ::::::::::|||:
 720 yArgAsnValSerAsnIleThrAlaAsnIleThrAlaThrAspAsnAlaL 737

2135 AAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGT 2184
      :::::   :::|||            ::::::      :::|||:::|||
 737 ysValAsnLeuGlyTyr......LysAsnGlyAspGluValCysValArg 751

2185 TCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACCGACGA 2234
      ||||||:::||||||   ::::::|||
 752 SerAspTyrThrGlyTyrValThrCys..................... 760

2235 TAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTCGATC 2284
                                    :::||||||      |
 761 ............................AsnThrGlyAsn......L 765

2285 TTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGC 2334
      ||:::|||   |||   ||||||      |||   :::   :::||||||
 765 euSerAspLysAla...LeuAsnSerPheGlyAlaThrLysIleAsnGly 780

2335 AATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCAC 2384
      |||::::::   |||
 781 AsnValAsnLeuAsn................................. 785

2385 CCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATC 2434
                                  ||||||      ::::::::::
 786 ........................GlnAsnAlaAlaLeuValLeuGlyL 794

2435 AAGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTTAAT 2484
      ::|||:::|||   |||:::   :::::::::|||||||:::      :::
 794 ysAlaAlaLeuTrpGlyGlnIleGlnGlyGlnGlyAsnSerArgValSer 810

2485 CTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAACGC 2534
      |||:::::::|||
 811 LeuAsnGlnHis..................................... 814

2535 TAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTAGCCG 2584

 814 ..................................................... 814

2585 ATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGC 2634

 814 ..................................................... 814

2635 GGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCC 2684
                                    |||:::|||   |||
 815 ............................SerLysTrpHisLeuTh 820

2685 GTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACAC 2734
      :::   :::::::::   |||||||:::|||   ::::::   |||  |
 820 rGlyAspSerGlnValHisAsnLeuSerLeuAlaAspSerHisIleHisL 837

2735 TCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCG 2784
      |||||:::|||         ||||||   :::|||
 837 euAsnAsnAla......SerAspAlaGlnSerAla.............. 846

2785 ACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATC 2834
```

```
 846 ..................................................... 846

2835 CGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAA 2884
                                    ::::::::::::||||||   :::|
 847 ........................AsnGlnTyrHisThrLeuLysIleA 855

2885 ACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTC 2934
      ||       |||:::|||:::|||   |||:::::::::::::::|||
 855 sn...HisLeuSerGlyAsnGlyHisPheHisTyrLeuThrHisLeuAla 870

2935 GGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTA 2984
                ::::|||||||:::   :::   ||||||:::::::|||   ||
 871 GluAsnLeuGlyAspLysValLeuValLysGluSerAlaSerGlyHisTy 887

2985 CACCTTGGCGGTCAACAATACCGGCAACGAACCTGCAAGCCTCGAACAAT 3034
      |   |||   |||::::::      :::||||||   :::   |||   |
 887 rGlnLeuHisValGlnAspLysThrGlyGluPro...AsnGlnGluGlyL 903

3035 TGACGGTAGTGGAAGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAAT 3084
      ||:::::   ::::::::::::   :::      |||   :::|||:::
 903 euAsnLeuPheAspAlaSerSerValArgAspArgSer...HisLeuSer 918

3085 TTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACT 3134
        :::|||   |||:::|||||||||   ||||||   ||||||   ::
 919 ValSerLeuAlaAsnHisHisValAspLeuGlyAlaLeuArgTyrThrIl 935

3135 CATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAG 3184
      :      :::::::|||      ||||||:::||||||      |||::::
 935 eLysThrGluAsnGlyIleThrArgLeuTyrAsnProTyrAlaGluAsnA 952

3185 AGCTTTCCGACAAACTCGGCAAGGCAGAAGCC.........AAAAAACAG 3225
      ::          |||   :::   :::   |||            :::|||
 952 rgArgArgValLysProAlaProSerProAlaThrAsnThrAlaSerGln 968

3226 GCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCG 3275
      |||:::   |||:::||||||                ||||||      ::
 969 AlaGlnThrAspSerAlaGln..............IleAlaLysProGl 980

3276 CGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCA 3324
      ::::   |||                ||||||   |||:::   ||||||
 980 nAsnIleVal..............ValAlaProProSerProGlnAla 991
```

```
seq_name: sp_bacteria:O30573

seq_documentation_block:
ID   O30573      PRELIMINARY;      PRT;   992 AA.
AC   O30573;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z3910;
RX   MEDLINE=98010345; PubMed=9350862;
RA   Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
RA   Achtman M.;
RT   "Clonal descent and microevolution of Neisseria meningitidis during 30
RT   years of epidemic spread.";
RL   Mol. Microbiol. 25:1047-1064(1997).
DR   EMBL; AF012205; AAC45788.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR002195; Dihydrooratase.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
DR   PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW   Protease.
```

```
FT   NON_TER       1      1
FT   NON_TER     992    992
SQ   SEQUENCE   992 AA;  109255 MW;  EF7CA563BC7D7ECB CRC64;


alignment_scores:
            Quality: 1368.00              Length:   1183
              Ratio:  2.060                Gaps:     36
 Percent Similarity:  56.128   Percent Identity: 31.192

alignment_block:
 US-09-303-518D-649 x O30573   ..

 Align seg 1/1  to: O30573  from: 1  to: 992

     94 CTGTCGTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGG 143
        |||:::::::::::::   |||   :::   |||:::   :::
      1 LeuAlaTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAs 17

    144 CATCAACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTG 193
        ::::::||||||   :::||||||||||||||||||||||||||||||||
     17 pValAspTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPheP 34

    194 CAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC 243
        |||||||||   |||:::::::|||   ||||||:::|||:::   :::
     34 heValGlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIle 50

    244 GGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCG 293
        |||:::::::::::::::::::|||||||||||||||||||||::::::
     51 GlyAsnAlaLeuSerAsnValProMetIleAspPheSerValAlaAspVa 67

    294 TAAC...GGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGG 340
        |||      :::   :::::::::      ||||||   |||||||||
     67 lAsnArgArgThrLeuThrValIleAspProGlnTyrAlaValSerValL 84

    341 CACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAAT 390
        |||      |||   :::         :::|||   ::::::   :::
     84 ysHisValLysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHis 100

    391 CCCGAT......CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAA 434
           |||      :::   :::      |||:::   |||   :::||::
    101 LeuAspValSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAs 117

    435 TTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGC............. 471
        :|||:::   :::   :::   |||   :::|||:::
    117 pTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuG 133

    472 ..GATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCT 519
          ||||||:::|||   ||||||::|||||||||||||:::::   |||
    133 luAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaPro 149

    520 GTTGAAATGACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAA 566
        :::      ||||||         |||   :::   |||   |||:::||
    150 IleAlaProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysAs 166

    567 TAATTACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGAT 616
        |:::::   :::   ||||||:::||||||||   |||:::
    166 nArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe....... 180

    617 CTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGT 666
                  |||   |||:::|||   :::      :::::::::
    181 ..........GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArg 193

    667 GCGTATTCTTGGCTCGTTGGTGGCAATACCTTT.............GC 701
        ||||||   :::   ::::::|||:::   :::             ::
    194 AlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValTh 210

    702 ACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTA 751
        ::::|||   :::      |||   ::::::::::|||::::::::::
    210 rSerAsnLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLysH 227

    752 AACATAGC.....................CCATATGGTTTTTTACCA 777
        ||||||                       :::   |||
```

```
    227 isHisSerLeuGluLysLeuLysGluValLeuSerGlnAsnAlaLeuThr 243

    778 ACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGC 827
        :::   :::   :::|||||||||||||||||||:::|||   ||||||
    244 AsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrAspLy 260

    828 CCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCT 877
         |||::::::::::|||::::::   |||:::   :::           |
    260 sGlnGluLysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGlyT 277

    878 ATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG...TTC 924
        ||      :::::::|||:::::::|||            :::|||   :::
    277 yr......GlnLysAsnSerTrpGln............GluTrpAsnIle 287

    925 TATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACG 974
        |||   :::   ||||||                              |||   ::
    288 TyrLysLysGluPheAlaAsp.....................GluIleLy 297

    975 TCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAA 1024
        :|||:::                        ||||||   ||||||   |||
    297 sGlnArg.....................AspAsnAlaGlyThrIleLysG 307

   1025 TCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGA 1074
            :::::::|||:::   ::::::                    |||
    307 lyTyrGlyGluHisHisTrpLysThr.................ThrGly 317

   1075 ACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGT 1124
        |||          |||      :::::::::||||||   :::   ::
    318 Thr............AsnSerHisIleGlySerThrAlaValArgLeuAl 330

   1125 TTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATG 1174
        :   :::      |||:::                          |||||||
    330 aGlyAsnGluArgGlyAla.....................AsnAsnG 339

   1175 GAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACC 1224
        ||:::|||::::::|||      |||::::::|||   |||:::|||
    339 lyGlnAsnValThrPhe......GluAsnAsnGlyThrLeuValLeuAsp 353

   1225 AGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTAC 1274
        :::||||||||||||||||||||||||||:::|||:::||||||:::||
    354 GlnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrTh 370

   1275 GGTCTCGCCTGAAAATAACGAA...ACTTGGCAAGGCGCGGGCGTTCATA 1321
        ||||::      ||||||      ||||||   |||||||||:::   :
    370 rValLysGlyAlaAsnAsnGlyIleThrTrpLeuGlyAlaGlyIleAspV 387

   1322 TCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGAC 1371
        ::::::::   :::   |||:::|||:::|||::::::      ::::||
    387 alAlaAspGlyLysLysValValTrpGlnValLysAsnProAsnGlyAsp 403

   1372 CGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGA 1421
        ||||||:::|||||||||||||||||||||:::::::::::::   |||
    404 ArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyVa 420

   1422 AAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGC 1471
         |||||||||:::::::::|||||||||||||||||||||||:::||||
    420 lAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnG 437

   1472 AGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTC 1521
        ||||||||         |||   ||||||||||:::::|||:::|||
    437 lnAlaAspAlaAspLysLysValGlnAlaPheSerGlnValGlyIleVal 453

   1522 AGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGA 1571
        ||||||||||||||::   ||||||:::   :::|||:::||||||::
    454 SerGlyArgGlyThrLeuValLeuAsnSerProAspGlnIleAsnProAs 470

   1572 CAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATT 1621
        ::::||||||||||||||||||||||||||||||||   ||||||::::
    470 nAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnA 487

   1622 CGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTC 1671
        ::|||::|||::::::|||::|||:::||||||||||||   ||||||
    487 spLeuThrPheGluHisIleArgAsnValAspGluGlyAlaArgIleVal 503
```

```
1672 AACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGC..... 1713
     |||||||||   |||:::   ||||||:::|||:::||||||
 504 AsnHisAsnThrAspArgAlaSerThrIleThrLeuThrGlyLysSerLe 520

1713 ..................................................... 1713

 520 uIleThrAlaProGlnAsnLeuSerValTyrGluIleArgAsnAspTyrA 537

1713 ..................................................... 1713

 537 spAspAspAspTyrTyrGlyTyrTyrSerTyrArgLysProIleProGln 553

1714 AATAAAGATATT.............................................. 1725
     :::|||||||:::
 554 GlyLysAspLeuTyrTyrLysAsnTyrArgTyrTyrAlaLeuLysSerGl 570

1726 ..........................GCTACAACCGGCAATAAC. 1743
                                :::   |||   ||||||
 570 yGlySerValAsnAlaProMetProGluAsnGlyGlnThrGluAsnAsnA 587

1743 ..................................................... 1743

 587 spTrpIleLeuMetGlySerThrGlnGluGluAlaLysLysAsnAlaMet 603

1744 AACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGA 1793
     |||     ::::::::::::      :::::::::|||:::||||||||
 604 AsnHisLysAsnAsnGlnArgIleSerGlyPheSerGlyPhePheGlyGl 620

1794 GAAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCG 1843
     |::::::            ||||||   |||||||||   ::::::
 620 uGluAsnGlyLysGlyHisAsnGlyAlaLeuAsnLeuAsnPheAsnGlyL 637

1844 CCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGC 1893
       :::   ::::::   :::|||||||:::|||||||||||||||||||||
 637 ysSerAlaGlnAsnArgPheLeuLeuThrGlyGlyThrAsnLeuAsnGly 653

1894 AACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAACACC 1943
     :::|||:::   |||:::|||:::::::::::::|||||||||||||||||
 654 LysIleSerValThrGlnGlyAsnValLeuLeuSerGlyArgProThrPr 670

1944 GCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGC.... 1989
     |||||||   :::   :::|||         :::|||||::::::
 670 oHisAlaArgAspPheValAsnLysSerSerAlaGlnLysAspAlaHisP 687

1990 ..ATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACA 2037
           ::::::|||:::|||:::::::::::|||||||||||||||||||
 687 heSerLysAsnAsnGluValValPheGluAspAspTrpIleAsnArgThr 703

2038 TTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCC.. 2085
     |||||||||   ::::::   ::::::   :::   :::      |||
 704 PheLysAlaThrGluIleThrValAsnGlnSerAlaSerPheSerSerGl 720

2086 .CGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCC 2134
      |||||||||:::::::::   ::::::          ::::::::::|||:
 720 yArgAsnValSerAsnIleThrAlaAsnIleThrAlaThrAspAsnAlaL 737

2135 AAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGT 2184
     :::::   :::|||                ::::::      :::|||:::|||
 737 ysValAsnLeuGlyTyr......LysAsnGlyAspGluValCysValArg 751

2185 TCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACCGACGA 2234
     ||||||:::||||||   :::::::|||         :::::::::   ||
 752 SerAspTyrThrGlyTyrValThrCysAsnThrGlyAsnLeuSer...As 767

2235 TAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTCGATC 2284
     ||||::::::   |||:::      |||:::|||:::||||||||||:::|
 767 pLysAlaLeuAsnSerPheGlyAlaThrGlnIleAsnGlyAsnValAsnL 784

2285 TTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGC 2334
     ||
 784 eu..................................................  784
```

```
2335 AATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCAC 2384

 784 ..................................................  784

2385 CCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATC 2434
          |||   |||   :::||||||                  :::::::
 785 ....AsnGlnAsnAlaAlaLeuVal..................LeuGlyL 794

2435 AAGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTTAAT 2484
     ::|||:::|||   |||:::   ::::::::::||||||:::      :::
 794 ysAlaAlaLeuTrpGlyGlnIleGlnGlyGlnGlyAsnSerArgValSer 810

2485 CTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAACGC 2534
     |||::::::|||
 811 LeuAsnGlnHis..................................... 814

2535 TAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTAGCCG 2584

 814 ..................................................  814

2585 ATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGC 2634

 814 ..................................................  814

2635 GGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCC 2684
                                        |||:::|||   |||
 815 ..............................SerLysTrpHisLeuTh 820

2685 GTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTACAC 2734
       :::   :::::::::   |||||||:::|||   ::::::   |||   |
 820 rGlyAspSerGlnValHisAsnLeuSerLeuAlaAspSerHisIleHisL 837

2735 TCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCG 2784
     ||||||:::|||          ||||||   :::|||
 837 euAsnAsnAla......SerAspAlaGlnSerAla.............. 846

2785 ACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATC 2834

 846 ..................................................  846

2835 CGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAA 2884
                                    :::::::::::::||||||   :::|
 847 ........................AsnGlnTyrHisThrLeuLysIleA 855

2885 ACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTC 2934
     ||       |||:::|||:::|||   |||:::::::::::::::|||
 855 sn...HisLeuSerGlyAsnGlyHisPheHisTyrLeuThrHisLeuAla 870

2935 GGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTA 2984
                ::::||||||::   :::   ||||||:::::|||   ||
 871 GluAsnLeuGlyAspLysValLeuValLysGluSerAlaSerGlyHisTy 887

2985 CACCTTGGCGGTCAACAATACCGGCAACGAACCTGCAAGCCTCGAACAAT 3034
     |   |||   |||::::::      :::||||||   :::   |||   |
 887 rGlnLeuHisValGlnAspLysThrGlyGluPro...AsnGlnGluGlyL 903

3035 TGACGGTAGTGGAAGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAAT 3084
     ||::::::   ::::::::::::   :::      |||   :::|||:::
 903 euAsnLeuPheAspAlaSerSerValArgAspArgSer...HisLeuSer 918

3085 TTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACT 3134
       :::|||   |||:::|||||||||   ||||||   ||||||   ::
 919 ValSerLeuAlaAsnHisHisValAspLeuGlyAlaLeuArgTyrThrIl 935

3135 CATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAG 3184
     :      ::::::|||      ||||||:::|||||      |||::::
 935 eLysThrGluAsnGlyIleThrArgLeuTyrAsnProTyrAlaGluAsnA 952

3185 AGCTTTCCGACAAACTCGGCAAGGCAGAAGCC.........AAAAAACAG 3225
     ::        |||   :::   :::   |||           :::|||
 952 rgArgArgValLysProAlaProSerProAlaThrAsnThrAlaSerGln 968

3226 GCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCG 3275
```

```
      |||:::   |||:::||||||                |||||     ::
    969 AlaGlnThrAspSerAlaGln.............IleAlaLysProGl 980

   3276 CGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCA 3324
        ::::   |||               ||||||   |||:::   ||||||
    980 nAsnIleVal.............ValAlaProProSerProGlnAla 991

seq_name: sp_bacteria:O30574

seq_documentation_block:
ID   O30574      PRELIMINARY;      PRT;   996 AA.
AC   O30574;
DT   01-JAN-1998 (TrEMBLrel. 05, Created)
DT   01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Z4024;
RX   MEDLINE=98010345; PubMed=9350862;
RA   Morelli G., Malorny B., Muller K., Seiler A., Wang J.F., del Valle J.,
RA   Achtman M.;
RT   "Clonal descent and microevolution of Neisseria meningitidis during 30
RT   years of epidemic spread.";
RL   Mol. Microbiol. 25:1047-1064(1997).
DR   EMBL; AF012206; AAC45789.2; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR002195; Dihydrooratase.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
DR   PROSITE; PS00482; DIHYDROOROTASE_1; UNKNOWN_1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     996    996
SQ   SEQUENCE   996 AA;  109688 MW;  B3ABDEFE54C337D9 CRC64;


alignment_scores:
            Quality: 1364.00            Length:   1188
              Ratio:  2.039              Gaps:     39
 Percent Similarity:  56.313   Percent Identity: 31.313

alignment_block:
 US-09-303-518D-649 x O30574   ..

 Align seg 1/1  to: O30574  from: 1  to: 996

      94 CTGTCGTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGG 143
         |||:::::::::::::   |||   :::   |||:::   :::
       1 LeuAlaTyrAlaLeuThrProTyrSerGluAlaAlaLeuValArgAspAs 17

     144 CATCAACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTG 193
         ::::::||||||   :::||||||||||||||||||||||||||||||
      17 pValAspTyrGlnIlePheArgAspPheAlaGluAsnLysGlyLysPheP 34

     194 CAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTC 243
          |||||||||   |||:::::::|||   ||||||:::|||:::   :::
      34 heValGlyAlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIle 50

     244 GGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCG 293
         |||:::::::::::::::::::|||||||||||||||||||||:::::
      51 GlyAsnAlaLeuSerAsnValProMetIleAspPheSerValAlaAspVa 67

     294 TAAC...GGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGG 340
          |||      :::   :::::::::      ||||||   |||||||||
      67 lAsnArgArgThrLeuThrValIleAspProGlnTyrAlaValSerValL 84

     341 CACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAAT 390
```

```
         |||      |||   :::         :::|||   ::::::   :::
      84 ysHisValLysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHis 100

     391 CCCGAT......CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAA 434
         |||      :::   :::      |||:::   |||   :::|||::
     101 LeuAspValSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAs 117

     435 TTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGC..............  471
         :|||:::   :::   :::   |||   :::|||:::
     117 pTyrGluProAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuG 133

     472 ..GATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCT 519
          ||||||:::|||   ||||||:::||||||||||||:::::   |||
     133 luAspTyrAsnMetAlaArgLeuAsnLysPheValThrGluValAlaPro 149

     520 GTTGAAATGACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAA 566
         :::      ||||||      |||   :::   |||   |||:::||
     150 IleAlaProThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysAs 166

     567 TAATTACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGAT 616
         |:::::   :::   ||||||:::||||||||||   |||:::
     166 nArgPheSerGluPheValArgValGlyAlaGlyThrGlnPhe.......  180

     617 CTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGT 666
                  |||   |||:::|||   :::      :::::::::
     181 ..........GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArg 193

     667 GCGTATTCTTGGCTCGTTGGTGGCAATACCTTT..............GC 701
         ||||||   :::   :::::::|||:::   :::              ::
     194 AlaTyrArgTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValTh 210

     702 ACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTA 751
         ::::|||   :::      |||   :::::::::|||:::::::::
     210 rSerAsnLeuAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLysH 227

     752 AACATAGCCCA.....................TATGGTTTTTTACCA 777
          |||||||||                      :::   |||
     227 isHisSerProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuThr 243

     778 ACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGC 827
         :::   :::   :::|||||||||||||||||||:::||||   ||||||
     244 AsnTyrAlaValLeuGlyAspSerGlySerProLeuPheAlaTyrAspLy 260

     828 CCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCT 877
          |||::::::::::|||:::::   |||:::   :::            |
     260 sGlnGluLysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGlyT 277

     878 ATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG...TTC 924
         ||      ::::::|||::::::|||              :::|||   :::
     277 yr......GlnLysAsnSerTrpGln............GluTrpAsnIle 287

     925 TATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACG 974
         |||   :::   ||||||                        |||   ::
     288 TyrLysLysGluPheAlaAsp.....................GluIleLy 297

     975 TCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAA 1024
         :|||:::                    ||||||   ||||||   |||
     297 sGlnArg.....................AspAsnAlaGlyThrIleLysG 307

    1025 TCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGA 1074
              ::::::|||:::   ::::::                   |||
     307 lyTyrGlyGluHisHisTrpLysThr..................ThrGly 317

    1075 ACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGT 1124
         |||            |||      :::::::::||||||
     318 Thr............AsnSerHisIleGlySerThrAla........... 326

    1125 TTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATG 1174
             :::   ||||||:::   ::::::              |||||||
     327 .ValArgLeuAlaGlyAsnGluArgAsp............AlaAsnAsnG 339

    1175 GAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACC 1224
         ||:::|||::::::|||      |||   :::|||   |||:::|||
```

```
 339 lyGlnAsnValThrPhe......GluAspAsnGlyThrLeuValLeuAsp 353

1225 AGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTAC 1274
     :::|||||||||||||||||||||||||||||:::|||:::||||||:::|| |
 354 GlnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrTh 370

1275 GGTCTCGCCTGAAAATAACGAA...ACTTGGCAAGGCGCGGGCGTTCATA 1321
     ||||:::      ||||||     ||||||   ||||||||||:::   :
 370 rValLysGlyAlaAsnAsnGlyIleThrTrpLeuGlyAlaGlyIleAspV 387

1322 TCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGAC 1371
     ::::::::   :::   |||:::|||:::|||:::::::      :::||| |
 387 alAlaAspGlyLysLysValValTrpGlnValLysAsnProAsnGlyAsp 403

1372 CGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGA 1421
     ||||||:::|||||||||||||||||||||||||||:::::::::::   |||
 404 ArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyVa 420

1422 AAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGC 1471
     |||||||||:::::::::|||||||||||||||||||||||:::||||:
 420 lAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnL 437

1472 AGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTC 1521
     ::||||||::::::   |||   ||||||||||||:::::|||:::||| |
 437 ysAlaAspSerAsnGlnLysValGlnAlaPheSerGlnValGlyIleVal 453

1522 AGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCGGA 1571
     ||||||||||||||::   ||||||:::::|||||||:::|||||||| |
 454 SerGlyArgGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAs 470

1572 CAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATT 1621
     |:::|||||||||||||||||||||||||||||||   ||||||:::: |
 470 pAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnA 487

1622 CGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTC 1671
     ::|||:::|||:::::|||:::|||:::|||||||||||||   |||||| |
 487 spLeuThrPheGluHisIleArgAsnValAspGluGlyAlaArgIleVal 503

1672 AACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGC..... 1713
     |||||||||   :::      ||||||:::|||:::||||||
 504 AsnHisAsnThrSerHisAlaSerThrIleThrLeuThrGlyLysSerLe 520

1713 ..................................................... 1713

 520 uIleThrAlaProGlnAsnLeuSerValTyrGluIleArgAsnAspTyrA 537

1713 ..................................................... 1713

 537 spAspAspAspTyrTyrGlyTyrTyrSerTyrArgLysProIleProGln 553

1714 AATAAAGATATT..................................... 1725
     :::|||||||:::
 554 GlyLysAspLeuTyrTyrLysAsnTyrArgTyrTyrAlaLeuLysSerGl 570

1726 ............................GCTACAACCGGCAATAAC. 1743
                                 ::::::|||   ||||||
 570 yGlySerValAsnAlaProMetProGluAsnGlyValThrGluAsnAsnA 587

1743 ..................................................... 1743

 587 spTrpValPheMetGlyTyrThrGlnGluGluAlaLysLysAsnAlaMet 603

1744 AACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGA 1793
     |||      ::::::::::::      ::::::::::|||:::|||||||| |
 604 AsnHisLysAsnAsnGlnArgIleSerGlyPheSerGlyPhePheGlyGl 620

1794 GAAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCG 1843
     |::::::      ||||||   |||||||||   ::::::
 620 uGluAsnGlyLysGlyHisAsnGlyAlaLeuAsnLeuAsnPheAsnGlyL 637

1844 CCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGC 1893
     :::   ::::::   :::||||||:::||||||:::|||||||||||||
 637 ysSerAlaGlnAsnArgPheLeuLeuThrGlyGlyAlaAsnLeuAsnGly 653

1894 AACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAACACC 1943
     :::|||:::   |||:::|||:::::::::::::|||||||||||||||||
 654 LysIleSerValThrGlnGlyAsnValLeuLeuSerGlyArgProThrPr 670

1944 GCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGC.... 1989
     |||||||   :::   :::|||          :::::::|||::::::
 670 oHisAlaArgAspPheValAsnLysSerSerAlaArgLysAspAlaHisP 687

1990 ..ATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACA 2037
           ::::::|||:::|||:::::::::|||||||||||||||||||
 687 heSerLysAsnAsnGluValValPheGluAspAspTrpIleAsnArgThr 703

2038 TTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCC.. 2085
     |||||||||   ::::::   ::::::   :::   :::      |||
 704 PheLysAlaAlaGluIleThrValAsnGlnSerAlaSerPheSerSerGl 720

2086 .CGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCC 2134
      ||||||||||:::   :::   ::::::          ::::::::|||:
 720 yArgAsnValSerAspIleThrAlaAsnIleThrAlaThrAspAsnAlaL 737

2135 AAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGT 2184
     :::::   :::|||           ::::::      :::|||:::|||
 737 ysValAsnLeuGlyTyr......LysAsnGlyAspGluValCysValArg 751

2185 TCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACCGACGA 2234
     ||||||:::||||||   ::::::|||
 752 SerAspTyrThrGlyTyrValThrCys..................... 760

2235 TAAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTCGATC 2284
                          :::||||||             :::|
 761 .....................AsnThrAsp.............AsnL 765

2285 TTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGC 2334
     ||:::|||   |||                              ||||||
 765 euSerAspLysAla.........................LeuAsn... 771

2335 AATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCAC 2384
                                        |||   :::||||||
 772 ..................................SerPheGlyAlaTh 776

2385 CCAA...AACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTA 2431
     ||||   |||||||||:::::|||      ||||||      :::::::
 776 rGlnIleAsnGlyAsnValAsnLeuSerGlnAsnAlaAlaLeuValLeuG 793

2432 ATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTT 2481
     :::::|||:::|||   |||:::   :::::::::||||||:::
 793 lyLysAlaAlaLeuTrpGlyGlnIleGlnGlyGlnGlyAsnSerArgVal 809

2482 AATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAA 2531
     :::|||:::::|||
 810 SerLeuAsnGlnHis................................. 814

2532 CGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTAG 2581

 814 ..................................................... 814

2582 CCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATC 2631

 814 ..................................................... 814

2632 AGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCT 2681
                                        |||:::|||   ||
 815 ..................................SerLysTrpHisLe 819

2682 GCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACCATTA 2731
     |   :::   :::::::::   ||||||:::|||   ::::::   |||
 819 uThrGlyAspSerGlnValHisAsnLeuSerLeuAlaAspSerHisIleH 836

2732 CACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAAACCGGCAGT 2781
       ||||||:::|||          ||||||   :::|||
 836 isLeuAsnAsnAla......SerAspAlaGlnSerAla............ 846
```

```
  2782 GCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATT 2831

   846 ..................................................... 846

  2832 ATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGG 2881
                                 :::::::::::::||||||   :
   847 ........................AsnLysTyrHisThrLeuLysI 854

  2882 TAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTC 2931
       ::|||      |||:::|||:::|||   |||:::::::::::::|||
   854 leAsn...HisLeuSerGlyAsnGlyHisPheHisTyrLeuThrHisLeu 869

  2932 TTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCAC 2981
                  :::||||||::::   :::   ||||||::::::|||
   870 AlaLysAsnLeuGlyAspLysValLeuValLysGluSerAlaSerGlyHi 886

  2982 TTACACCTTGGCGGTCAACAATACCGGCAACGAACCTGCAAGCCTCGAAC 3031
       |||   |||   |||:::::::      :::||||||   :::   |||
   886 sTyrGlnLeuHisValGlnAspLysThrGlyGluPro...AsnGlnGluG 902

  3032 AATTGACGGTAGTGGAAGGAAAAGACAACAAACCGCTGTCCGAAAACCTT 3081
         |||::::::   ::::::::::::   :::      |||   :::|||
   902 lyLeuAsnLeuPheAspAlaSerSerValGlnAspArgSer...ArgLeu 917

  3082 AATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCA 3131
       :::   :::|||   |||:::||||||||   ||||||   ||||||
   918 SerValSerLeuAlaAsnAsnHisValAspLeuGlyAlaLeuArgTyrTh 934

  3132 ACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAAC 3181
         :::      ::::::|||      ||||||:::||||||      |||:
   934 rIleLysThrGluAsnGlyIleThrArgLeuTyrAsnProTyrAlaGluA 951

  3182 AAGAGCTTTCCGACAAACTCGGC.....................AAGGCA 3210
       :::::         |||   :::                        :::
   951 snArgArgArgValLysProAlaProSerProAlaThrAsnThrAlaSer 967

  3211 GAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCT 3260
       :::|||:::|||   ::::::   |||:::||||||
   968 GlnAlaGlnLysAlaThrGlnThrAspGlyAlaGln..............  979

  3261 GATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAAC 3310
        ||||||      ::::::   |||                  ||||||   |
   980 .IleAlaLysProGlnAsnIleVal..............ValAlaProP 991

  3311 CGGCCCGGCAGGCA 3324
       ||:::   ||||||
   991 roSerProGlnAla 995
```

seq_name: sp_bacteria:Q47692

```
seq_documentation_block:
ID   Q47692      PRELIMINARY;      PRT;  1377 AA.
AC   Q47692;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   TSH.
GN   TSH.
OS   Escherichia coli.
OG   Plasmid pAPEC-1.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CHI7122;
RX   MEDLINE=94178945; PubMed=8132344;
RA   Provence D.L., Curtiss R. III;
RT   "Isolation and characterization of a gene involved in hemagglutination
RT   by an avian pathogenic Escherichia coli strain.";
RL   Infect. Immun. 62:1369-1380(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CHI7122;
RX   MEDLINE=20316042; PubMed=10858231;
RA   Dozois C.M., Dho-Moulin M., Bree A., Fairbrother J.M., Desautels C.,
RA   Curtiss R. III;
RT   "Relationship between the Tsh autotransporter and pathogenicity of
RT   avian Escherichia coli and localization and analysis of the Tsh
RT   genetic region.";
RL   Infect. Immun. 68:4145-4154(2000).
DR   EMBL; AF218073; AAA24698.1; -.
DR   MEROPS; S06.003; -.
DR   InterPro; IPR000710; IgA_S6.
DR   InterPro; IPR001254; Trypsin.
DR   Pfam; PF02395; IGA1; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
KW   Hydrolase; Plasmid; Serine protease.
SQ   SEQUENCE   1377 AA;  148226 MW;  237423644D9AE012 CRC64;
```

```
alignment_scores:
            Quality: 1111.50              Length:   1573
              Ratio:  1.388                 Gaps:     62
 Percent Similarity:  50.922   Percent Identity: 25.366

alignment_block:
 US-09-303-518D-649 x Q47692   ..

 Align seg 1/1  to: Q47692  from: 1  to: 1377

     90 ATGCCTGTCGTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATT 139
        :::||||||:::      |||:::   ||||||:::||||||  :::
     42 IleProValLeu...PheSerAlaGlySerLeuAlaGlyThr.ValAsnA 57

    140 TCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAG 189
           ::::::|||||   :::||||||||||||||||||||||||
     57 snGluLeuGlyTyrGlnLeuPheArgAspPheAlaGluAsnLysGlyMet 73

    190 TTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTT 239
        |||      ||||||   :::|||   :::|||||||||:::||||||::
     74 PheArgProGlyAlaThrAsnIleAlaIleTyrAsnLysGlnGlyGluPh 90

    240 GGTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTG. 288
        :||||||   ::::::   ||||||   |||   ||||||||||:::|||
     90 eValGly...ThrLeuAspLysAlaAlaMetProAspPheSerAlaValA 106

    289 ..TCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGC 336
          |||:::   |||||||||:::|||::::::   |||||||||:::|||
    106 spSerGluIleGlyValAlaThrLeuIleAsnProGlnTyrIleAlaSer 122

    337 GTGGCACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAG 386
        |||   |||||||||||||||:::||||||:::||||||   :::||||::
    123 ValLysHisAsnGlyGlyTyrThrAsnValSerPheGly...AspGlyGl 138

    387 AAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATT 436
        :|||                  |||:::||||||   |||||||||
    138 uAsn.................ArgTyrAsnIleValAspArgAsnAsnA 149

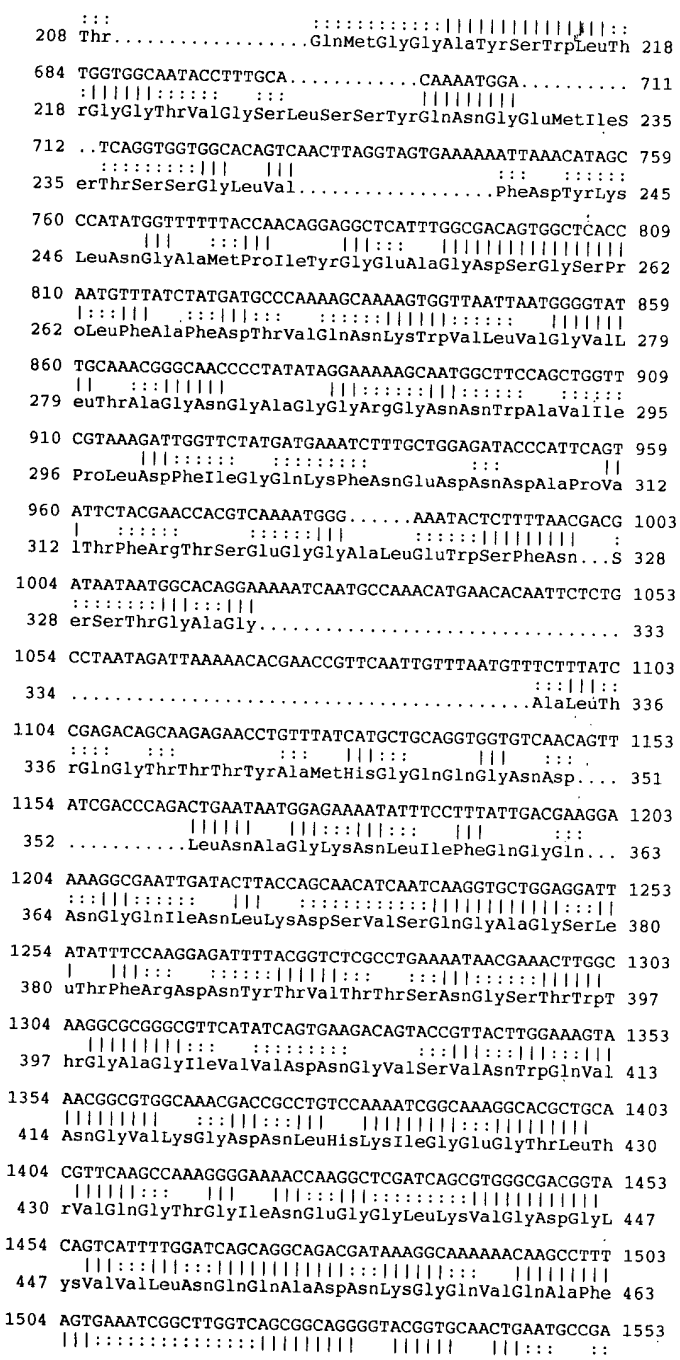
    437 ATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCG 486
           :::                                |||:::|||   |||
    149 laProSer.......................LeuAspPheHisAlaPro 157

    487 CGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTA 536
        ||||||   |||:::||||||::::::   |||:::   :::||||:::
    158 ArgLeuAspLysLeuValThrGluValAlaProThrAlaValThrAlaGl 174

    537 T...ATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTG 583
                  :::      |||:::|||::::::::::||||||
    174 nGlyAlaValAlaGlyAlaTyrLeuAspLysGluArgTyrProValPheT 191

    584 TTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAGCCC 633
          |||:::|||:::|||   ||||||   ::::::::::::   :::
    191 yrArgLeuGlySerGlyThrGlnTyrIleLysAspSerAsnGlyGlnLeu 207

    634 AATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGT 683
```


208 Thr.....GlnMetGlyGlyAlaTyrSerTrpLeuTh 218
684 TGGTGGCAATACCTTTGCA.....CAAAATGGA..... 711
218 rGlyGlyThrValGlySerLeuSerSerTyrGlnAsnGlyGluMetIleS 235
712 ..TCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGC 759
235 erThrSerSerGlyLeuVal.....PheAspTyrLys 245
760 CCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACC 809
246 LeuAsnGlyAlaMetProIleTyrGlyGluAlaGlyAspSerGlySerPr 262
810 AATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTAT 859
262 oLeuPheAlaPheAspThrValGlnAsnLysTrpValLeuValGlyValL 279
860 TGCAAACGGGCAACCCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTT 909
279 euThrAlaGlyAsnGlyAlaGlyGlyArgGlyAsnAsnTrpAlaValIle 295
910 CGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGT 959
296 ProLeuAspPheIleGlyGlnLysPheAsnGluAspAsnAspAlaProVa 312
960 ATTCTACGAACCACGTCAAAATGGG.....AAATACTCTTTTAACGACG 1003
312 lThrPheArgThrSerGluGlyGlyAlaLeuGluTrpSerPheAsn...S 328
1004 ATAATAATGGCACAGGAAAAATCAATGCCAAACATGAACACAATTCTCTG 1053
328 erSerThrGlyAlaGly..... 333
1054 CCTAATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATC 1103
334 .....AlaLeuTh 336
1104 CGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTT 1153
336 rGlnGlyThrThrThrTyrAlaMetHisGlyGlnGlnGlyAsnAsp.... 351
1154 ATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGA 1203
352 .....LeuAsnAlaGlyLysAsnLeuIlePheGlnGlyGln... 363
1204 AAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATT 1253
364 AsnGlyGlnIleAsnLeuLysAspSerValSerGlnGlyAlaGlySerLe 380
1254 ATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGC 1303
380 uThrPheArgAspAsnTyrThrValThrThrSerAsnGlySerThrTrpT 397
1304 AAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTA 1353
397 hrGlyAlaGlyIleValValAspAsnGlyValSerValAsnTrpGlnVal 413
1354 AACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCA 1403
414 AsnGlyValLysGlyAspAsnLeuHisLysIleGlyGluGlyThrLeuTh 430
1404 CGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTA 1453
430 rValGlnGlyThrGlyIleAsnGluGlyGlyLeuLysValGlyAspGlyL 447
1454 CAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTT 1503
447 ysValValLeuAsnGlnGlnAlaAspAsnLysGlyGlnValGlnAlaPhe 463
1504 AGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGA 1553
464 SerSerValAsnIleAlaSerGlyArgProThrValValLeuThrAspGl 480
1554 TAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTT 1603
480 uArgGlnValAsnProAspThrValSerTrpGlyTyrArgGlyGlyThrL 497
1604 TGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGAT 1653
497 euAspValAsnGlyAsnSerLeuThrPheHisGlnLeuLysAlaAlaAsp 513
1654 GAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTAC 1703
514 TyrGlyAlaValLeuAlaAsn...AsnValAspLysArgAlaThrIleTh 529
1704 CATTACAGGCAATAAAGATATTGCTACAACCGGCAATAACAACAGCTTGG 1753
529 rLeu.....AspTyrAlaLeuArgAlaAsp..... 537
1754 ATAGCAAAAAAGAAATTGCCTACAACGGTTGG..... 1785
538 .....LysValAlaLeuAsnGlyTrpSerGluSerGlyLysGly 550
1785 1785
551 ThrAlaGlyAsnLeuTyrLysTyrAsnAsnProTyrThrAsnThrThrAs 567
1786 ....TTTGGCGAGAAAGATACGACC..... 1806
567 pTyrPheIleLeuLysGlnSerThrTyrGlyTyrProThrAspGlnS 584
1806 1806
584 erSerAsnAlaThrTrpGluPheValGlyHisSerGlnGlyAspAlaGln 600
1807 .....AA 1808
601 LysLeuValAlaAspArgPheAsnThrAlaGlyTyrLeuPheHisGlyGl 617
1809 AACGAACGGGCGGCTCAACCTTGTTTACCAG...CCCGCCGCAGAAGACC 1855
617 nLeuLysGlyAsnLeuAsnValAspAsnArgLeuProGluGlyValThrG 634
1856 GCACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAA 1905
634 lyAlaLeuValMetAspGlyAlaAlaAspIleSerGlyThrPheThrGln 650
1906 ACAAACGGCAAACTGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAA 1955
651 GluAsnGlyArgLeuThrLeuGlnGlyHisProValIleHisAlaTyrAs 667
1956 TCATTTAAAC.....GACCAT..... 1971
667 nThrGlnSerValAlaAspLysLeuAlaAlaSerGlyAspHisSerValL 684
1972 .....TGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAA 2004
684 euThrGlnProThrSerPheSerGlnGlu..... 693
2005 ATCGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCGGAAAACTT 2054
694 .....AspTrpGluAsnArgSerPheThrPheAspArgLe 705
2055 CCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAAGTGA 2104
705 uSerLeuLysAsnThrAspPheGlyLeuGlyArgAsnAlaThr..... 719
2105 AAGGCGATTGGCATTTGAGCAATCACGCCCAAGCA...GTTTTTGGTGTC 2151
720 .....LeuAsnThrThrIleGlnAlaAspAsnSerServal 731
2152 GCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCT 2201
732 ThrLeuGlyAspSerArgValPheIleAspLysAsnAspGlyGlnGlyTh 748


2202 GACAAATTGTGTCGAAAAA.....ACCATTACCGACGATAAAGTGA 2242
748 rAlaPheThrLeuGluGluGlyThrSerValAlaThrLysAsp..... 762
2243 TTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTCGATCTTGCCGAT 2292
763 .....AlaAspLysSerValPheAsnGlyThrValAsnLeuAspAsn 776
2293 CACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAG 2342
777 GlnSerValLeuAsnIleAsnAspIle.....PheAsnGlyGlyIleGl 791
2343 TGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACG 2392
791 nAlaAsnAsnSerThrValAsnIleSerSerAspSerAla..... 804
2393 GCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACA 2442
805 .....ValLeuGlyAsnSerThrLeuThrSerThrAlaLeuAsn 817
2443 TTAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTT..... 2481
818 LeuAsnLysGlyAlaAsnAlaLeuAlaSerGlnSerPheValSerAspGl 834
2482 .....AATCTAAGCGACCACGCCGTACAA..... 2505
834 yProValAsnIleSerAspAlaAlaLeuSerLeuAsnSerArgProAspG 851
2506 .....AACGGCAGTCTGACG 2520
851 luValSerHisThrLeuLeuProValTyrAspTyrAlaGlySerTrpAsn 867
2521 CTTTCCGGC...AACGCTAAGGCAAACGTAAGC...CATTCCGCACTCAA 2564
868 LeuLysGlyAspAspAlaArgLeuAsnValGlyProTyrSerMetLeuSe 884
2565 CGGTAATGTCTCCCTAGCCGATAAG..... 2589
884 rGlyAsnIleAsnValGlnAspLysGlyThrValThrLeuGlyGlyGluG 901
2590 .....GCAGTA 2595
901 lyGluLeuSerProAspLeuThrLeuGlnAsnGlnMetLeuTyrSerLeu 917
2596 TTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGA 2645
918 PheAsnGlyTyrArgAsnIleTrpSerGlySerLeuAsn...AlaProAs 933
2646 TACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGG 2695
933 pAlaThrValSerMetThrAspThrGlnTrpSerMetAsnGlyAsnSerT 950
2696 AATTAGGCAATTTAAACCTTGACAACGCCACCATTACACTCAATTCCGCC 2745
950 hrAlaGlyAsnMetLysLeuAsnArgThrIleValGlyPheAsnGlyGly 966
2746 TATCGCCACGATGCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCC 2795
966 966
2796 GCGCCGCCGTTCGCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGC 2845
966 966
2846 CAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTG 2895
967 .....ThrSerProPheThrThrLeuThrThrAspAsnLeuAsp 979
2896 AACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAG 2945
980 AlaValGlnSerAlaPheValMetArgThrAspLeu.....AsnLysAl 994
2946 CGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGG 2995
994 aAspLysLeuValIleAsnLysSerAla..... 1003
2996 TCAACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTG 3045
1003 1003
3046 GAAGGAAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCA 3095
1004 ThrGlyHisAspAsn.....SerIleTrpValAsnPhe..... 1014
3096 AAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAG 3145
1014 1014
3146 ACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGAC 3195
1014 1014
3196 AAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCA 3245
1015 .....LeuLysLysProSerAsnLysAsp..... 1022
3246 AAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAG 3295
1023 .ThrLeuAspIleProLeu.....ValSerAlaProG 1033
3296 AAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGC... 3342
1033 luAlaThrAlaAspAsnLeuPheArgAlaSerThrArgValValGlyPhe 1049
3343 .....ATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCA 3377
1050 SerAspValThrProIleLeuSerValArgLysGluAspGlyLysLysGl 1066
3378 GGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGC 3427
1066 uTrpValLeuAspGlyTyrGlnValAlaArgAsnAspGlyGlnGlyLysA 1083
3428 CGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCG 3477
1083 laAlaAlaThr..... 1086
3478 CAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCAGCCG 3527
1086 1086
3528 TTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTT 3577
1087 .PheMetHisIleSerTyrAsnAsnPheIleThrGluValAsnAsnLeuA 1103
3578 TCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGACCGCCGCAAC 3627
1103 snLysArgMetGlyAspLeuArgAspIleAsnGlyGlu.....Ala 1116
3628 GCCGTTTGGACA.....AGCGGCATCCGGGAC..... 3654
1117 GlyThrTrpValArgLeuLeuAsnGlySerGlySerAlaAspGlyGlyPh 1133
3655 .ACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCG 3703
1133 eThrAspHisTyr.....T 1138
3704 ACCTGCGCCAAATCGGTATGCAGAAA.....AACCTCGGCAGCGGGCGC 3747
1138 hrLeuLeuGlnMetGlyAlaAspArgLysHisGluLeuGlySerMetAsp 1154
3748 GTC.....GGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGA 3791
1155 LeuPheThrGlyValMetAlaThrTyrThrAspThrAspAlaSerAlaAs 1171
3792 CGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGC 3841

```
                 |          |||::::::          :::   |||
 1171 pLeuTyrSerGlyLysThrLysSerTrpGlyGlyGly..... 1183

 3842 AATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTT 3891
                          ||||||            |||::::::   |||
 1184 .................PheTyr............AlaSerGlyLeuPhe 1190

 3892 AGCAGCGGCAGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGT 3941
        ||||||:::      |||   |||:::      |||:::::::
 1191 ArgSerGlyAlaTyrPheAspValIleAlaLysTyrIleHisAsn..... 1205

 3942 GCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCG 3991
                      :::   :::|||       :::|||:::|||   |
 1206 ................GluAsnLysTyrAspLeuAsnPheAlaGlyAlaG 1217

 3992 GCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGAT 4041
      ||   :::         :::  :::   |||         ::::::
 1217 lyLysGlnAsnPheArgSerHisSerLeuTyrAlaGlyAlaGluValGly 1233

 4042 TACCGCTACGAA......AACGTCAATATCGCCACCCCC........... 4074
      |||||||||:::      ::::::::::::::   |||
 1234 TyrArgTyrHisLeuThrAspThrThrPheValGluProGlnAlaGluLe 1250

 4075 ..................GGCCTTGCATTCAACCGCTACCGCGCGGGCA 4105
                         |||   :::|||||||          :::|||:
 1250 uValTrpGlyArgLeuGlnGlyGlnThrPheAsnTrpAsnAspSerGlyM 1267

 4106 TTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCT 4155
      ::      |||   |||::::::          :::   |||::::::|||
 1267 et......AspValSerMetArg.........ArgAsnSerValAsnPro 1278

 4156 TATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACG 4205
                ::::::   |||   ::::::||||||||:::   :::
 1279 Leu......ValGlyArgThrGlyValValSerGlyLysThrPheSerGl 1293

 4206 CGTCAATACCGCCGTATTGGCTCAG..............GATTTCGGCA 4240
           :::   ::::::   |||:::              :::|||
 1293 yLysAspTrpSerLeuThrAlaArgAlaGlyLeuHisTyrGluPheAspL 1310

 4241 AAACCCGCAGTGCGGAA 4257
        |||   ||||||:::
 1310 euThrAspSerAlaAsp 1315

seq_name: sp_bacteria:O88093

seq_documentation_block:
ID   O88093      PRELIMINARY;      PRT;  1377 AA.
AC   O88093;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HAEMOGLOBIN PROTEASE PRECURSOR.
GN   HBP.
OS   Escherichia coli.
OG   Plasmid pColV-K30.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98416173; PubMed=9743528;
RA   Otto B.R., van Dooren S.J.M., Nuijens J.H., Luirink J., Oudega B.;
RT   "Characterization of a hemoglobin protease secreted by the pathogenic
RT   Escherichia coli strain EB1.";
RL   J. Exp. Med. 188:1091-1103(1998).
DR   EMBL; AJ223631; CAA11507.1; -.
DR   MEROPS; S06.003; -.
DR   InterPro; IPR000710; IgA_S6.
DR   InterPro; IPR001254; Trypsin.
DR   Pfam; PF02395; IGA1; 1.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
KW   Hydrolase; Plasmid; Protease; Serine protease; Signal.
FT   SIGNAL       1     52       POTENTIAL.

SQ   SEQUENCE   1377 AA;  148256 MW;  BB16D898EDAC0416 CRC64;

alignment_scores:
           Quality: 1107.50                Length:   1573
             Ratio:  1.383                  Gaps:     62
 Percent Similarity:  50.922   Percent Identity: 25.302

alignment_block:
 US-09-303-518D-649 x O88093   ..

 Align seg 1/1  to: O88093  from: 1  to: 1377

      90 ATGCCTGTCGTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATT 139
         :::|||||||:::       |||:::   ||||||:::||||||| :::
      42 IleProValLeu...PheSerAlaGlySerLeuAlaGlyThr.ValAsnA 57

     140 TCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAG 189
                 ::::::||||||   :::|||||||||||||||||||||||||
      57 snGluLeuGlyTyrGlnLeuPheArgAspPheAlaGluAsnLysGlyMet 73

     190 TTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTT 239
         |||      ||||||   :::|||   :::|||||||||:::||||||::
      74 PheArgProGlyAlaThrAsnIleAlaIleTyrAsnLysGlnGlyGluPh 90

     240 GGTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTG. 288
         :||||||   ::::::  ||||||   |||   |||||||||:::|||
      90 eValGly...ThrLeuAspLysAlaAlaMetProAspPheSerAlaValA 106

     289 ..TCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGC 336
           |||:::   |||||||||:::|||::::::   |||||||||:::|||
     106 spSerGluIleGlyValAlaThrLeuIleAsnProGlnTyrIleAlaSer 122

     337 GTGGCACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAG 386
         |||   |||||||||||||||:::|||||||:::||||||   :::|||::
     123 ValLysHisAsnGlyGlyTyrThrAsnValSerPheGly...AspGlyGl 138

     387 AAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATT 436
         :|||                  |||:::||||||   |||||||||
     138 uAsn.................ArgTyrAsnIleValAspArgAsnAsnA 149

     437 ATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCG 486
           :::                                |||:::|||   |||
     149 laProSer.....................LeuAspPheHisAlaPro 157

     487 CGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTA 536
         ||||||   |||:::||||||::::::   |||:::   :::|||:::
     158 ArgLeuAspLysLeuValThrGluValAlaProThrAlaValThrAlaGl 174

     537 T...ATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTG 583
                 :::       |||:::|||:::::::::||||||
     174 nGlyAlaValAlaGlyAlaTyrLeuAspLysGluArgTyrProValPheT 191

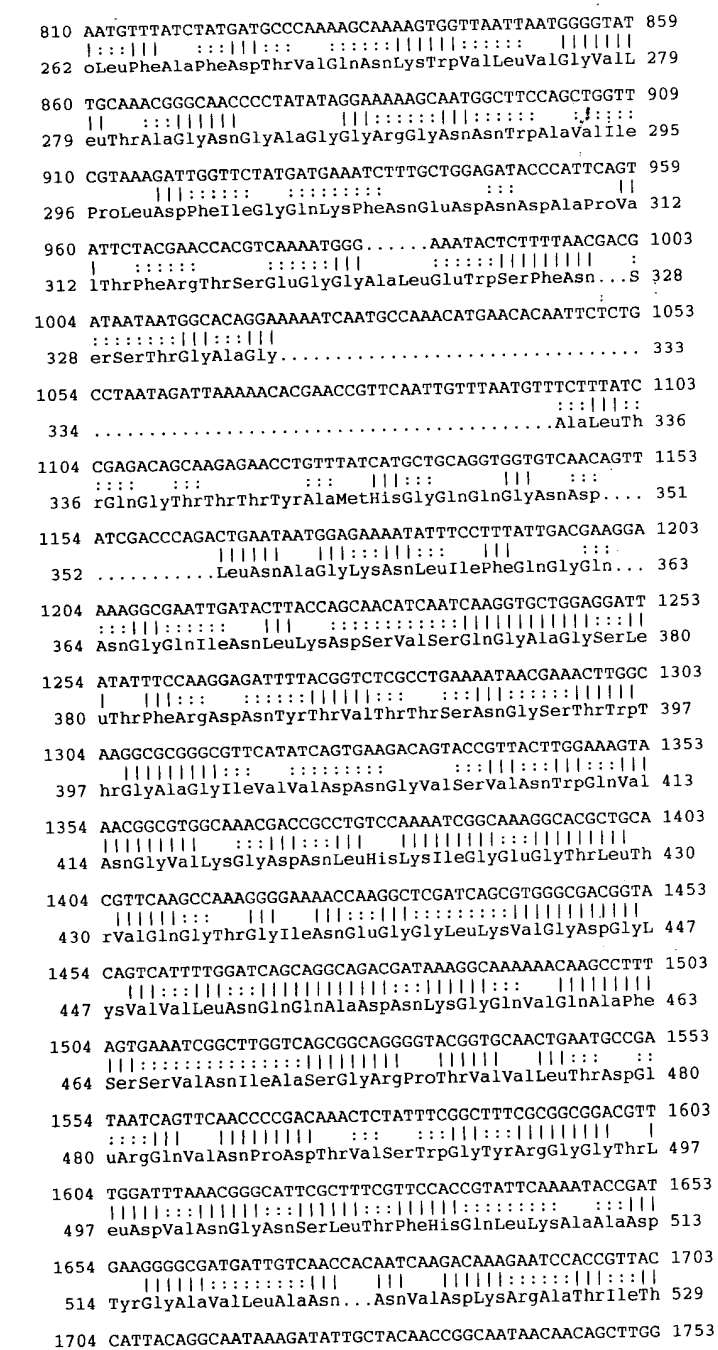
     584 TTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAGCCC 633
           |||:::|||:::|||   ||||||   ::::::::::::   :::
     191 yrArgLeuGlySerGlyThrGlnTyrIleLysAspSerAsnGlyGlnLeu 207

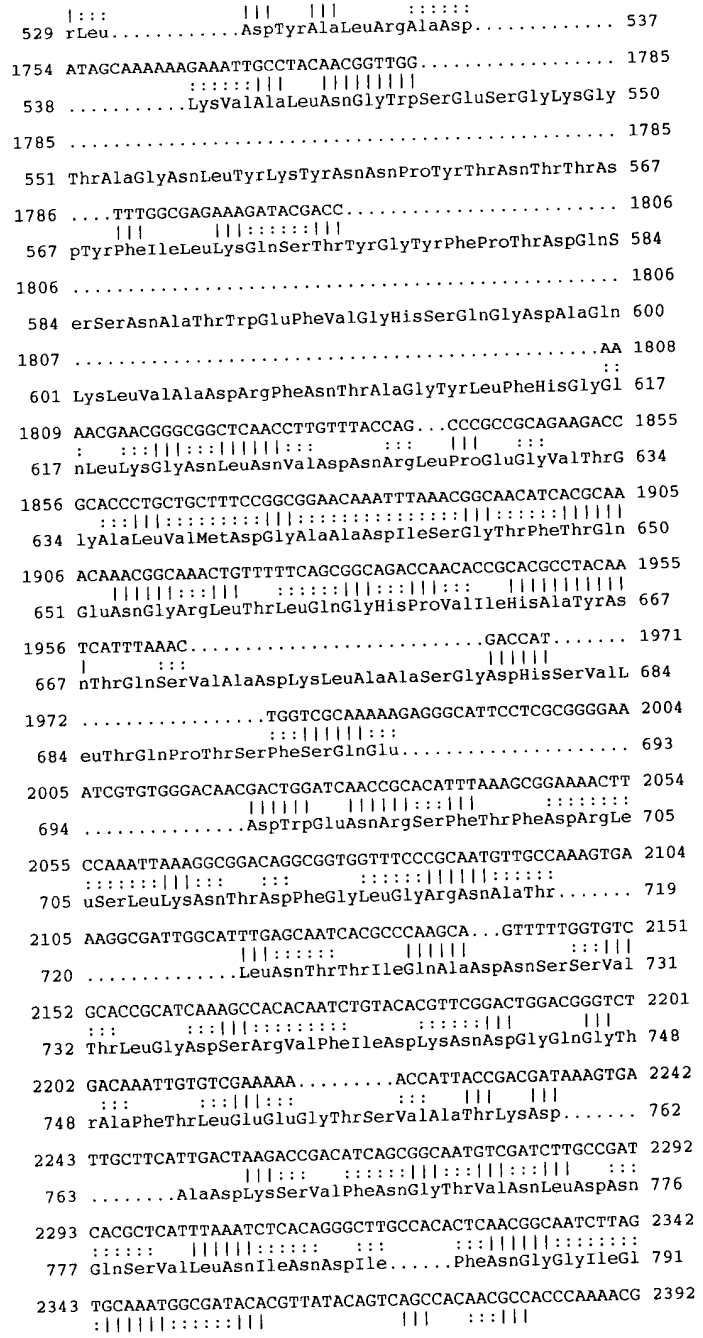
     634 AATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGT 683
         ::::::                 :::::::::|||||||||||||||::
     208 ThrLys.................MetGlyGlyAlaTyrSerTrpLeuTh 218

     684 TGGTGGCAATACCTTTGCA............CAAAATGGA......... 711
         :||||||::::::   :::            |||||||||
     218 rGlyGlyThrValGlySerLeuSerSerTyrGlnAsnGlyGluMetIleS 235

     712 ..TCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGC 759
           :::::::::|||   |||                  :::   ::::::
     235 erThrSerSerGlyLeuVal.................PheAspTyrLys 245

     760 CCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACC 809
               |||   :::|||       |||:::   ||||||||||||||||
     246 LeuAsnGlyAlaMetProIleTyrGlyGluAlaGlyAspSerGlySerPr 262
```







```
3992 GCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGAT 4041
1217 lyLysGlnAsnPheArgSerHisSerLeuTyrAlaGlyAlaGluValGly 1233

4042 TACCGCTACGAA......AACGTCAATATCGCCACCCCC.......... 4074
1234 TyrArgTyrHisLeuThrAspThrThrPheValGluProGlnAlaGluLe 1250

4075 ..................GGCCTTGCATTCAACCGCTACCGCGCGGGCA 4105
1250 uValTrpGlyArgLeuGlnGlyGlnThrPheAsnTrpAsnAspSerGlyM 1267

4106 TTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCT 4155
1267 et......AspValSerMetArg.........ArgAsnSerValAsnPro 1278

4156 TATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACG 4205
1279 Leu......ValGlyArgThrGlyValValSerGlyLysThrPheSerGl 1293

4206 CGTCAATACCGCCGTATTGGCTCAG.............GATTTCGGCA 4240
1293 yLysAspTrpSerLeuThrAlaArgAlaGlyLeuHisTyrGluPheAspL 1310

4241 AAACCCGCAGTGCGGAA 4257
1310 euThrAspSerAlaAsp 1315
```

seq_name: sp_bacteria:Q54151

seq_documentation_block:

```
ID   Q54151      PRELIMINARY;      PRT;  1372 AA.
AC   Q54151;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-JUL-1997 (TrEMBLrel. 04, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   HEMAGGLUTININ.
GN   SHE OR PIC.
OS   Shigella flexneri, and
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Shigella.
OX   NCBI_TaxID=623, 562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   SPECIES=S.flexneri; STRAIN=SEROTYPE 2A, AND M4243A;
RA   Noriega F.R.;
RL   Submitted (MAY-1997) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   SPECIES=E.coli; STRAIN=042;
RA   Henderson I.R., Czeczulin J.R., Nataro J.P.;
RT   "A second autotransporter protease from enteroaggregative Escherichia
RT   coli.";
RL   Submitted (OCT-1998) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U35656; AAB58244.1; -.
DR   EMBL; AF097644; AAD23953.1; -.
DR   MEROPS; S06.005; -.
DR   MEROPS; S06.UPW; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
KW   Signal; Protease.
SQ   SEQUENCE   1372 AA;  146449 MW;  68CE71254EAC26EA CRC64;
```

```
alignment_scores:
            Quality: 1074.00            Length:   1570
              Ratio:  1.395               Gaps:     65
 Percent Similarity:  49.045   Percent Identity: 25.924

alignment_block:
 US-09-303-518D-649 x Q54151   ..
```

Align seq 1/1 to: Q54151 from: 1 to: 1372

```
 85 GCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACAC 134
 44 AlaIleCysLeuCysTyrSerGlnIleSerGln......AlaGlyIleVa 58

135 TTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAATAAAG 184
 58 lArgSerAspIleAlaTyrGlnIleTyrArgAspPheAlaGluAsnLysG 75

185 GCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGG 234
 75 lyLeuPheValProGlyAlaAsnAspIleProValTyrAspLysAspGly 91

235 GAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGT 284
 92 LysLeuValGlyArg...LeuGlyLysAlaProMetAlaAspPheSerSe 107

285 GGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGA 334
107 rValSerSerAsnGlyValAlaThrLeuValSerProGlnTyrIleValS 124

335 GCGTGGCACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGA 384
124 erValLysHisAsnGlyGlyTyrArgSerValSerPheGly...AsnGly 139

385 AGAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAA 434
140 LysAsn.................ThrTyrSerLeuValAspArgAsnAs 150

435 TTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGC 484
150 n.....................HisPro...SerIleAspPheHisAlaP 159

485 CGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGT 534
159 roArgLeuAsnLysLeuValThrGluValIleProSerAlaValThrSer 175

535 TATATGGATGGGCGGAAA.........TATATCGATCAAAATAATTACCC 575
176 ......GluGlyThrLysAlaAsnAlaTyrLysTyrThrGluArgTyrTh 190

576 TGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAG 625
190 rAlaPheTyrArgValGlySerGlyThrGlnTyrThrLys...AspLysA 206

626 ATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTCT 675
206 sp..............GlyAsnLeuValLysValAlaGlyGlyTyrAla 217

676 TGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCAC 725
218 PheLysThrGlyGlyThrThr.......................GlyVa 226

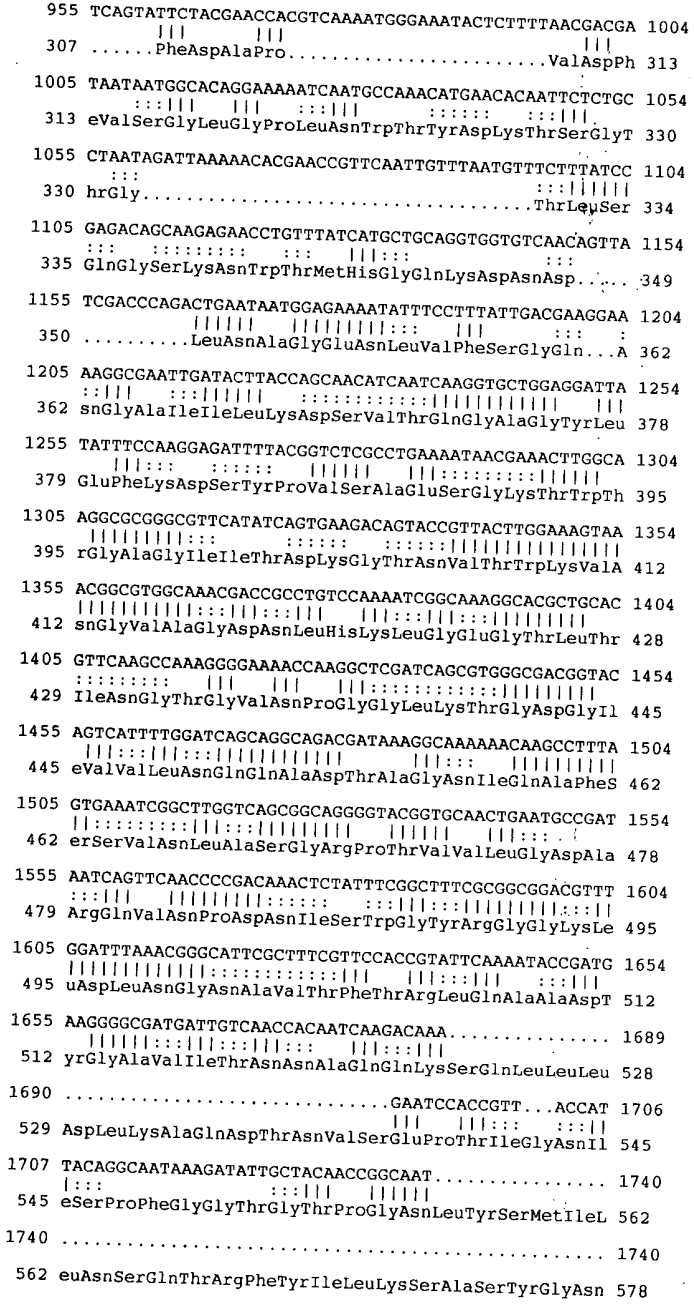
726 AGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCA............. 762
226 lProLeuIleSerAspAlaThrIleValSerAsnProGlyGlnThrTyrA 243

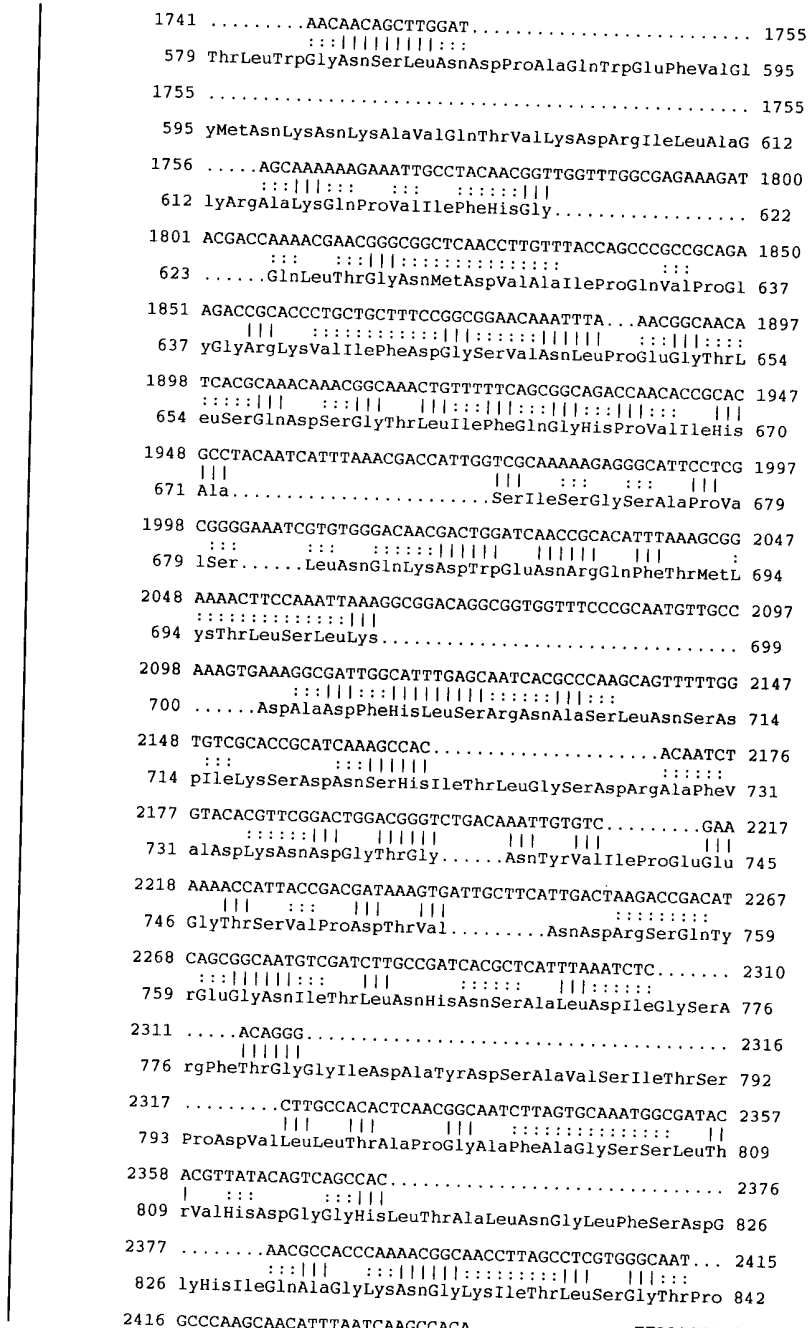
763 ........TATGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGGC 804
243 snProValAsnGlyProLeuProAspTyrGlyAlaProGlyAspSerGly 259

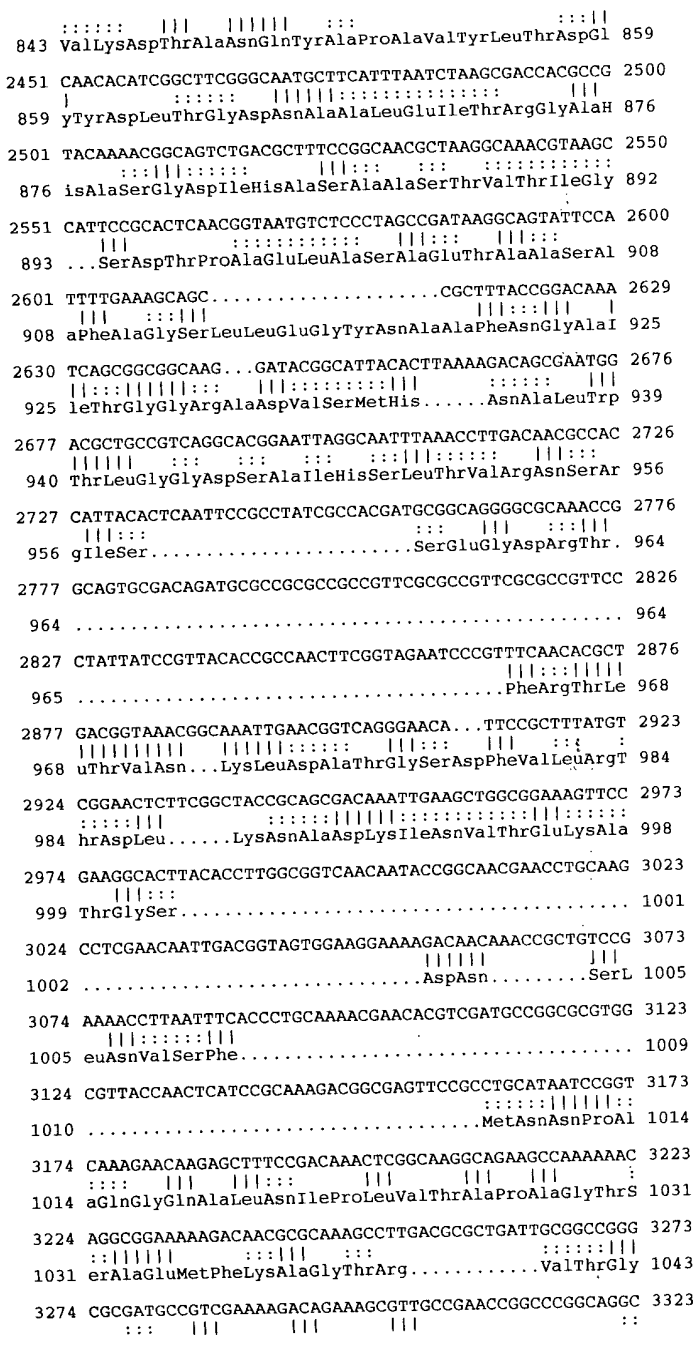
805 TCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGG 854
260 SerProLeuPheAlaTyrAspLysGlnGlnLysLysTrpValIleValAl 276

855 GGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGC 904
276 aValLeuArgAlaTyrAlaGlyIleAsnGlyAlaThrAsnTrpTrpAsnV 293

905 TGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCAT 954
293 alIleProThrAsp...TyrLeuAsnGlnValMetGlnAspAsp...... 306
```


955 TCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGA 1004
307 .....PheAspAlaPro.....ValAspPh 313
1005 TAATAATGGCACAGGAAAAATCAATGCCAAACATGAACACAATTCTCTGC 1054
313 eValSerGlyLeuGlyProLeuAsnTrpThrTyrAspLysThrSerGlyT 330
1055 CTAATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCC 1104
330 hrGly.....ThrLeuSer 334
1105 GAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTA 1154
335 GlnGlySerLysAsnTrpThrMetHisGlyGlnLysAspAsnAsp..... 349
1155 TCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAA 1204
350 .....LeuAsnAlaGlyGluAsnLeuValPheSerGlyGln...A 362
1205 AAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTA 1254
362 snGlyAlaIleIleLeuLysAspSerValThrGlnGlyAlaGlyTyrLeu 378
1255 TATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCA 1304
379 GluPheLysAspSerTyrProValSerAlaGluSerGlyLysThrTrpTh 395
1305 AGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAA 1354
395 rGlyAlaGlyIleIleThrAspLysGlyThrAsnValThrTrpLysValA 412
1355 ACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCAC 1404
412 snGlyValAlaGlyAspAsnLeuHisLysLeuGlyGluGlyThrLeuThr 428
1405 GTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTAC 1454
429 IleAsnGlyThrGlyValAsnProGlyGlyLeuLysThrGlyAspGlyIl 445
1455 AGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTA 1504
445 eValValLeuAsnGlnGlnAlaAspThrAlaGlyAsnIleGlnAlaPheS 462
1505 GTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGAT 1554
462 erSerValAsnLeuAlaSerGlyArgProThrValValLeuGlyAspAla 478
1555 AATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTT 1604
479 ArgGlnValAsnProAspAsnIleSerTrpGlyTyrArgGlyGlyLysLe 495
1605 GGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATG 1654
495 uAspLeuAsnGlyAsnAlaValThrPheThrArgLeuGlnAlaAlaAspT 512
1655 AAGGGGCGATGATTGTCAACCACAATCAAGACAAA..... 1689
512 yrGlyAlaValIleThrAsnAsnAlaGlnGlnLysSerGlnLeuLeuLeu 528
1690 .....GAATCCACCGTT...ACCAT 1706
529 AspLeuLysAlaGlnAspThrAsnValSerGluProThrIleGlyAsnIl 545
1707 TACAGGCAATAAAGATATTGCTACAACCGGCAAT..... 1740
545 eSerProPheGlyGlyThrGlyThrProGlyAsnLeuTyrSerMetIleL 562
1740 1740
562 euAsnSerGlnThrArgPheTyrIleLeuLysSerAlaSerTyrGlyAsn 578


1741 .....AACAACAGCTTGGAT..... 1755
579 ThrLeuTrpGlyAsnSerLeuAsnAspProAlaGlnTrpGluPheValGl 595
1755 1755
595 yMetAsnLysAsnLysAlaValGlnThrValLysAspArgIleLeuAlaG 612
1756 .....AGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGAT 1800
612 lyArgAlaLysGlnProValIlePheHisGly..... 622
1801 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGA 1850
623 .....GlnLeuThrGlyAsnMetAspValAlaIleProGlnValProGl 637
1851 AGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTA...AACGGCAACA 1897
637 yGlyArgLysValIlePheAspGlySerValAsnLeuProGluGlyThrL 654
1898 TCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAACACCGCAC 1947
654 euSerGlnAspSerGlyThrLeuIlePheGlnGlyHisProValIleHis 670
1948 GCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCG 1997
671 Ala.....SerIleSerGlySerAlaProVa 679
1998 CGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCGG 2047
679 lSer.....LeuAsnGlnLysAspTrpGluAsnArgGlnPheThrMetL 694
2048 AAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCC 2097
694 ysThrLeuSerLeuLys..... 699
2098 AAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGG 2147
700 .....AspAlaAspPheHisLeuSerArgAsnAlaSerLeuAsnSerAs 714
2148 TGTCGCACCGCATCAAAGCCAC.....ACAATCT 2176
714 pIleLysSerAspAsnSerHisIleThrLeuGlySerAspArgAlaPheV 731
2177 GTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTC.....GAA 2217
731 alAspLysAsnAspGlyThrGly.....AsnTyrValIleProGluGlu 745
2218 AAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACAT 2267
746 GlyThrSerValProAspThrVal.....AsnAspArgSerGlnTy 759
2268 CAGCGGCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTC..... 2310
759 rGluGlyAsnIleThrLeuAsnHisAsnSerAlaLeuAspIleGlySerA 776
2311 .....ACAGGG..... 2316
776 rgPheThrGlyGlyIleAspAlaTyrAspSerAlaValSerIleThrSer 792
2317 .....CTTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGATAC 2357
793 ProAspValLeuLeuThrAlaProGlyAlaPheAlaGlySerSerLeuTh 809
2358 ACGTTATACAGTCAGCCAC..... 2376
809 rValHisAspGlyGlyHisLeuThrAlaLeuAsnGlyLeuPheSerAspG 826
2377 .....AACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAAT... 2415
826 lyHisIleGlnAlaGlyLysAsnGlyLysIleThrLeuSerGlyThrPro 842
2416 GCCCAAGCAACATTTAATCAAGCCACA.....TTAAACGG 2450


843 ValLysAspThrAlaAsnGlnTyrAlaProAlaValTyrLeuThrAspGl 859
2451 CAACACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCG 2500
859 yTyrAspLeuThrGlyAspAsnAlaAlaLeuGluIleThrArgGlyAlaH 876
2501 TACAAAACGGCAGTCTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAGC 2550
876 isAlaSerGlyAspIleHisAlaSerAlaAlaSerThrValThrIleGly 892
2551 CATTCCGCACTCAACGGTAATGTCTCCCTAGCCGATAAGGCAGTATTCCA 2600
893 ...SerAspThrProAlaGluLeuAlaSerAlaGluThrAlaAlaSerAl 908
2601 TTTTGAAAGCAGC.....CGCTTTACCGGACAAA 2629
908 aPheAlaGlySerLeuLeuGluGlyTyrAsnAlaAlaPheAsnGlyAlaI 925
2630 TCAGCGGCGGCAAG...GATACGGCATTACACTTAAAAGACAGCGAATGG 2676
925 leThrGlyGlyArgAlaAspValSerMetHis......AsnAlaLeuTrp 939
2677 ACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCAC 2726
940 ThrLeuGlyGlyAspSerAlaIleHisSerLeuThrValArgAsnSerAr 956
2727 CATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAAACCG 2776
956 gIleSer.....SerGluGlyAspArgThr. 964
2777 GCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCGTTCC 2826
964 964
2827 CTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACACGCT 2876
965 .....PheArgThrLe 968
2877 GACGGTAAACGGCAAATTGAACGGTCAGGGAACA...TTCCGCTTTATGT 2923
968 uThrValAsn...LysLeuAspAlaThrGlySerAspPheValLeuArgT 984
2924 CGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTCC 2973
984 hrAspLeu......LysAsnAlaAspLysIleAsnValThrGluLysAla 998
2974 GAAGGCACTTACACCTTGGCGGTCAACAATACCGGCAACGAACCTGCAAG 3023
999 ThrGlySer..... 1001
3024 CCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACAAACCGCTGTCCG 3073
1002 .....AspAsn.....SerL 1005
3074 AAAACCTTAATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGG 3123
1005 euAsnValSerPhe..... 1009
3124 CGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGT 3173
1010 .....MetAsnAsnProAl 1014
3174 CAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAAC 3223
1014 aGlnGlyGlnAlaLeuAsnIleProLeuValThrAlaProAlaGlyThrS 1031
3224 AGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGG 3273
1031 erAlaGluMetPheLysAlaGlyThrArg.....ValThrGly 1043
3274 CGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGC 3323
1044 PheSerArgValThrProThrLeuHisVal.....AspThrSe 1056
3324 AGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGG 3373
1056 rGlyGlyAsnThrLysTrpIleLeuAspGly.....P 1067
3374 TGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACC 3423
1067 heLysAlaGluAlaAspLysAlaAlaAlaAlaLysAlaAspSer..... 1081
3424 CGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTT 3473
1081 1081
3474 GCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCA 3523
1081 1081
3524 GCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGC 3573
1082 .....PheMetAsnAlaGlyTyrLysAsnPheMetThrGluValAsnAsn 1096
3574 GTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGACCGCCG 3623
1097 Leu.....AsnLysAr 1100
3624 CAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGC 3673
1100 g.....MetGlyAspLeuArgAspThr.....AsnG 1109
3674 AAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATG 3723
1109 lyAspAlaGlyAlaTrpAlaArg.....Ile 1117
3724 CAGAAAAACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCG 3773
1118 MetSerGlyAlaGlySerAlaAspGlyGlyTyrSerAspAsnTyrThrHi 1134
3774 GACCGAAAACACCTTCGAC.....GACGGCATCGGC.... 3804
1134 sValGlnValGlyPheAspLysLysHisGluLeuAspGlyValAspLeuP 1151
3805 .....AACTCGGCACGGCTTGCCCACGGC 3828
1151 heThrGlyValThrMetThrTyrThrAspSerSerAlaAspSerHisAla 1167
3829 .....GCCGTTTTCGGGCAATACGGCATCGACAGGTT 3860
1168 PheSerGlyLysThrLysSerValGlyGly..... 1177
3861 CTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTCAG 3910
1178 .....GlyLeuTyrAlaSerAlaLeuPheGluSerGlyAlaTyrIleA 1192
3911 ACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTCAG 3960
1192 spLeuIleGly..... 1195
3961 GCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATCGG 4010
1195 1195
4011 CGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGTCA 4060
1196 .....LysTyrIleHisHisAspAsnAspTyrThrGlyAsnPheAlaS 1210
4061 ATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAG 4110
1210 erLeuGlyThrLysHisTyrAsnThrHisSerTrpTyrAlaGlyAlaGlu 1226
4111 GCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTT 4160
1227 ThrGlyTyrArgTyrHisLeuThrGluAspThrPheIleGluProGlnAl 1243

```
     4161 GAGCCTGTCCTATACCGATGCCGCTTCGGGCAAA..............  4194
          :::|||   |||      |||:::||||||||||
     1243 aGluLeuValTyr...GlyAlaValSerGlyLysThrPheArgTrpLysA 1259

     4195 ...................GTCCGAACACGCGTCAATACCGCCGTATTG 4224
                             ::::::::::|||      :::   ::::::
     1259 spGlyAspMetAspLeuSerMetLysAsnArgAspPheSerProLeuVal 1275

     4225 GCTCAG.........GATTTCGGCAAAACCCGCAGTGCG...GAATGGGG 4262
          ::::::         ::::::|||||||||   |||:::   :::|||::
     1276 GlyArgThrGlyValGluLeuGlyLysThrPheSerGlyLysAspTrpSe 1292

     4263 CGTAAACGCC 4272
          :|||:::|||
     1292 rValThrAla 1295

seq_name: sp_bacteria:Q9AL58

seq_documentation_block:
ID   Q9AL58      PRELIMINARY;      PRT;  1373 AA.
AC   Q9AL58;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PIC.
GN   PIC.
OS   Shigella flexneri 2a.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Shigella.
OX   NCBI_TaxID=42897;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YSH6000T;
RX   MEDLINE=20231776; PubMed=10768931;
RA   Al-Hasani K., Henderson I.R., Sakellaris H., Rajakumar K., Grant T.,
RA   Nataro J.P., Robins-Browne R., Adler B.;
RT   "The sigA gene which is borne on the she pathogenicity island of
RT   Shigella flexneri 2a encodes an exported cytopathic protease involved
RT   in intestinal fluid accumulation.";
RL   Infect. Immun. 68:2457-2463(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=YSH6000T;
RX   MEDLINE=21105956; PubMed=11162180;
RA   Al-Hasani K., Rajakumar K., Bulach D., Robins-Browne R., Adler B.,
RA   Sakellaris H.;
RT   "Genetic organization of the she pathogenicity island in Shigella
RT   flexneri 2a.";
RL   Microb. Pathog. 30:1-8(2001).
DR   EMBL; AF200692; AAK00464.1; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
SQ   SEQUENCE   1373 AA;  146548 MW;  45C016CDAA577763 CRC64;


alignment_scores:
           Quality: 1074.00              Length:   1570
             Ratio:  1.395                Gaps:     65
 Percent Similarity:  49.045   Percent Identity: 25.924

alignment_block:
 US-09-303-518D-649 x Q9AL58   ..

 Align seg 1/1  to: Q9AL58  from: 1  to: 1373

      85 GCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACAC 134
         ||||||||||||   ::::::   :::   |||      ||||||   ::
      45 AlaIleCysLeuCysTyrSerGlnIleSerGln......AlaGlyIleVa 59

     135 TTATTTCGGCATCAACTACCAATACTATCGCGACTTTGCCGAAAATAAAG 184
         :         |||   ||||||   |||||||||||||||||||||||||
      59 lArgSerAspIleAlaTyrGlnIleTyrArgAspPheAlaGluAsnLysG 76

     185 GCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGG 234
         ||   |||:::   ||||||::||||||   |||||||:::|||   |||
      76 lyLeuPheValProGlyAlaAsnAspIleProValTyrAspLysAspGly 92

     235 GAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGT 284
         :::|||||||||:::   :::   ||||||||||||   |||||||||
      93 LysLeuValGlyArg...LeuGlyLysAlaProMetAlaAspPheSerSe 108

     285 GGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGA 334
         ||||||   ||||||||||||:::||||||:::   ||||||||||||
     108 rValSerSerAsnGlyValAlaThrLeuValSerProGlnTyrIleValS 125

     335 GCGTGGCACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGA 384
         |||||   |||||||||||||||:::::|||:::||||||   :::|||
     125 erValLysHisAsnGlyGlyTyrArgSerValSerPheGly...AsnGly 140

     385 AGAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAA 434
         :::|||                  ||||||:::::::|||   ||||||||
     141 LysAsn.................ThrTyrSerLeuValAspArgAsnAs 151

     435 TTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGC 484
         |                   ||||||   :::   |||:::|||   |
     151 n.....................HisPro...SerIleAspPheHisAlaP 160

     485 CGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGT 534
         ||||||||:::|||:::|||||:::::   |||      :::|||||||
     160 roArgLeuAsnLysLeuValThrGluValIleProSerAlaValThrSer 176

     535 TATATGGATGGGCGGAAA.........TATATCGATCAAAATAATTACCC 575
               :::|||   |||         |||         :::::::|||
     177 ......GluGlyThrLysAlaAsnAlaTyrLysTyrThrGluArgTyrTh 191

     576 TGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAG 625
                 |||:::|||:::|||   ||||||   :::   ||||:::|
     191 rAlaPheTyrArgValGlySerGlyThrGlnTyrThrLys...AspLysA 207

     626 ATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTCT 675
         ||                 :::         :::|||:::::|||:::
     207 sp..............GlyAsnLeuValLysValAlaGlyGlyTyrAla 218

     676 TGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCAC 725
         :::   :::||||||:::|||                        |||::
     219 PheLysThrGlyGlyThrThr.......................GlyVa 227

     726 AGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCA............. 762
         :      :::::::::       |||     :::|||
     227 lProLeuIleSerAspAlaThrIleValSerAsnProGlyGlnThrTyrA 244

     763 ........TATGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGGC 804
                 |||   ||||||      |||:::   ||||||||||||
     244 snProValAsnGlyProLeuProAspTyrGlyAlaProGlyAspSerGly 260

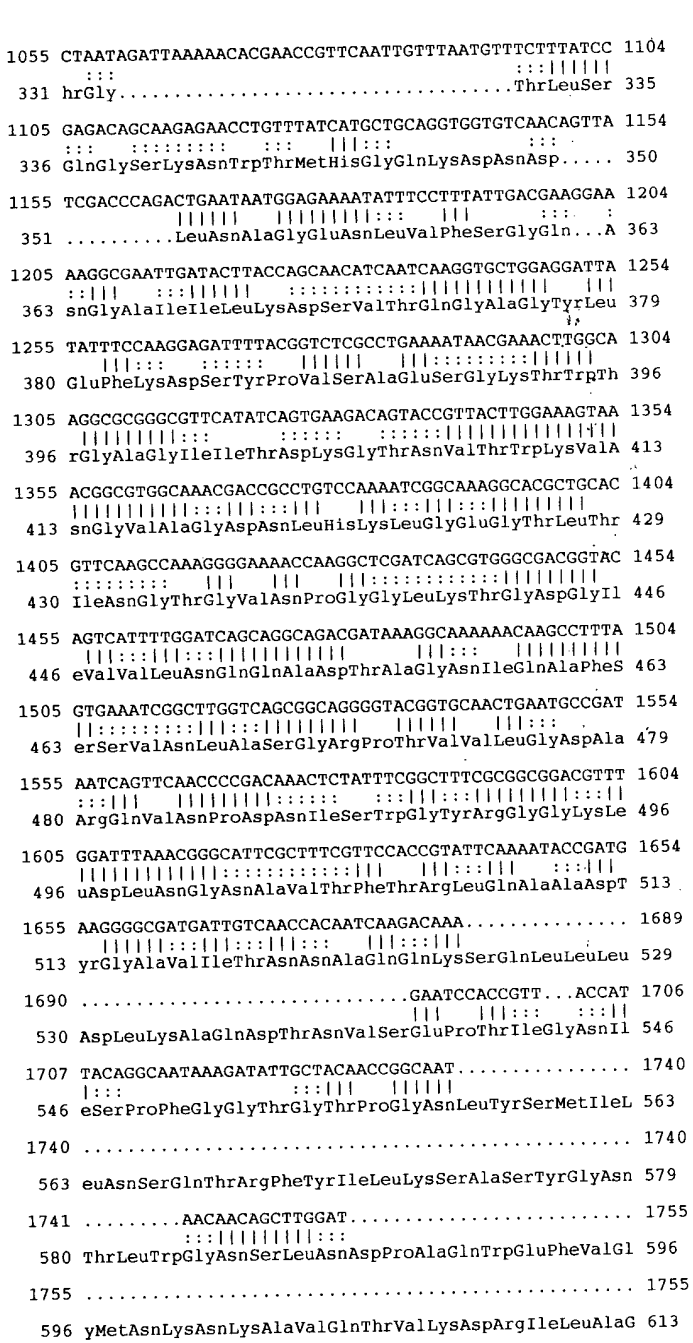
     805 TCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGG 854
         ||||||:::|||   ||||||   |||:::::|||||:::|||   ::
     261 SerProLeuPheAlaTyrAspLysGlnGlnLysLysTrpValIleValAl 277

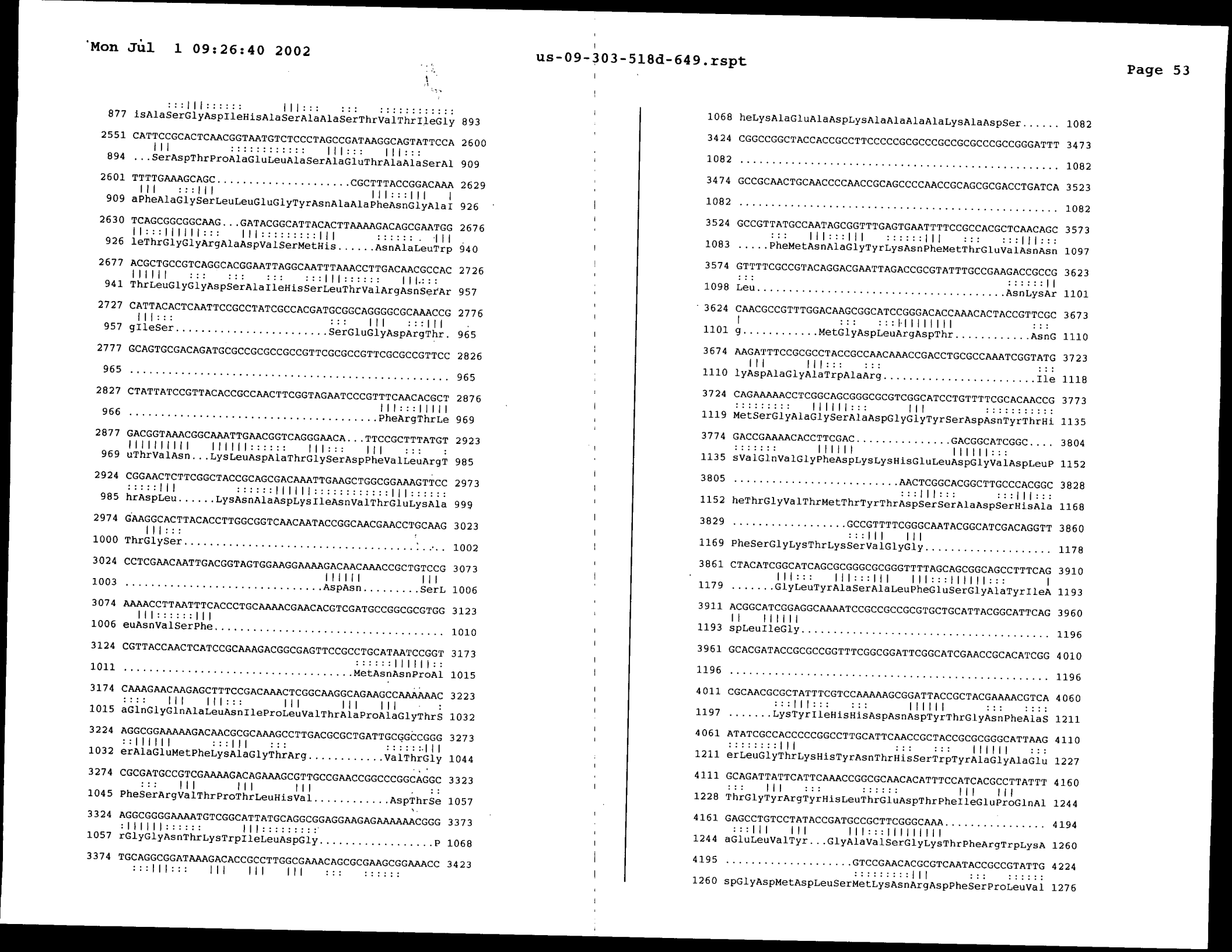
     855 GGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGC 904
         :||||||::::::               |||   :::|||   :::::::
     277 aValLeuArgAlaTyrAlaGlyIleAsnGlyAlaThrAsnTrpTrpAsnV 294

     905 TGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCAT 954
         :::::      |||   :::   ::::::::::::      |||
     294 alIleProThrAsp...TyrLeuAsnGlnValMetGlnAspAsp...... 307

     955 TCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGA 1004
               |||      |||                          |||
     308 ......PheAspAlaPro.......................ValAspPh 314

    1005 TAATAATGGCACAGGAAAAATCAATGCCAAACATGAACACAATTCTCTGC 1054
         :::|||   |||   :::|||      ::::::   :::|||
     314 eValSerGlyLeuGlyProLeuAsnTrpThrTyrAspLysThrSerGlyT 331
```


1055 CTAATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCC 1104
331 hrGly.....ThrLeuSer 335
1105 GAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTA 1154
336 GlnGlySerLysAsnTrpThrMetHisGlyGlnLysAspAsnAsp..... 350
1155 TCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAA 1204
351 .....LeuAsnAlaGlyGluAsnLeuValPheSerGlyGln...A 363
1205 AAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTA 1254
363 snGlyAlaIleIleLeuLysAspSerValThrGlnGlyAlaGlyTyrLeu 379
1255 TATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCA 1304
380 GluPheLysAspSerTyrProValSerAlaGluSerGlyLysThrTrpTh 396
1305 AGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAA 1354
396 rGlyAlaGlyIleIleThrAspLysGlyThrAsnValThrTrpLysValA 413
1355 ACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCAC 1404
413 snGlyValAlaGlyAspAsnLeuHisLysLeuGlyGluGlyThrLeuThr 429
1405 GTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTAC 1454
430 IleAsnGlyThrGlyValAsnProGlyGlyLeuLysThrGlyAspGlyIl 446
1455 AGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTA 1504
446 eValValLeuAsnGlnGlnAlaAspThrAlaGlyAsnIleGlnAlaPheS 463
1505 GTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGAT 1554
463 erSerValAsnLeuAlaSerGlyArgProThrValValLeuGlyAspAla 479
1555 AATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTT 1604
480 ArgGlnValAsnProAspAsnIleSerTrpGlyTyrArgGlyGlyLysLe 496
1605 GGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATG 1654
496 uAspLeuAsnGlyAsnAlaValThrPheThrArgLeuGlnAlaAlaAspT 513
1655 AAGGGGCGATGATTGTCAACCACAATCAAGACAAA..... 1689
513 yrGlyAlaValIleThrAsnAsnAlaGlnGlnLysSerGlnLeuLeuLeu 529
1690 .....GAATCCACCGTT...ACCAT 1706
530 AspLeuLysAlaGlnAspThrAsnValSerGluProThrIleGlyAsnIl 546
1707 TACAGGCAATAAAGATATTGCTACAACCGGCAAT..... 1740
546 eSerProPheGlyGlyThrGlyThrProGlyAsnLeuTyrSerMetIleL 563
1740 1740
563 euAsnSerGlnThrArgPheTyrIleLeuLysSerAlaSerTyrGlyAsn 579
1741 .....AACAACAGCTTGGAT..... 1755
580 ThrLeuTrpGlyAsnSerLeuAsnAspProAlaGlnTrpGluPheValGl 596
1755 1755
596 yMetAsnLysAsnLysAlaValGlnThrValLysAspArgIleLeuAlaG 613
1756 .....AGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGAT 1800
613 lyArgAlaLysGlnProValIlePheHisGly..... 623
1801 ACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGA 1850
624 .....GlnLeuThrGlyAsnMetAspValAlaIleProGlnValProGl 638
1851 AGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTA...AACGGCAACA 1897
638 yGlyArgLysValIlePheAspGlySerValAsnLeuProGluGlyThrL 655
1898 TCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAACACCGCAC 1947
655 euSerGlnAspSerGlyThrLeuIlePheGlnGlyHisProValIleHis 671
1948 GCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCG 1997
672 Ala.....SerIleSerGlySerAlaProVa 680
1998 CGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCGG 2047
680 lSer.....LeuAsnGlnLysAspTrpGluAsnArgGlnPheThrMetL 695
2048 AAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCC 2097
695 ysThrLeuSerLeuLys..... 700
2098 AAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGG 2147
701 .....AspAlaAspPheHisLeuSerArgAsnAlaSerLeuAsnSerAs 715
2148 TGTCGCACCGCATCAAAGCCAC.....ACAATCT 2176
715 pIleLysSerAspAsnSerHisIleThrLeuGlySerAspArgAlaPheV 732
2177 GTACACGTTCGGACTGGACGGGTCTGACAAATTGTGTC.....GAA 2217
732 alAspLysAsnAspGlyThrGly.....AsnTyrValIleProGluGlu 746
2218 AAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACCGACAT 2267
747 GlyThrSerValProAspThrVal.....AsnAspArgSerGlnTy 760
2268 CAGCGGCAATGTCGATCTTGCCGATCACGCTCATTTAAATCTC..... 2310
760 rGluGlyAsnIleThrLeuAsnHisAsnSerAlaLeuAspIleGlySerA 777
2311 .....ACAGGG..... 2316
777 rgPheThrGlyGlyIleAspAlaTyrAspSerAlaValSerIleThrSer 793
2317 .....CTTGCCACACTCAACGGCAATCTTAGTGCAAATGGCGATAC 2357
794 ProAspValLeuLeuThrAlaProGlyAlaPheAlaGlySerSerLeuTh 810
2358 ACGTTATACAGTCAGCCAC..... 2376
810 rValHisAspGlyGlyHisLeuThrAlaLeuAsnGlyLeuPheSerAspG 827
2377 .....AACGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAAT... 2415
827 lyHisIleGlnAlaGlyLysAsnGlyLysIleThrLeuSerGlyThrPro 843
2416 GCCCAAGCAACATTTAATCAAGCCACA.....TTAAACGG 2450
844 ValLysAspThrAlaAsnGlnTyrAlaProAlaValTyrLeuThrAspGl 860
2451 CAACACATCGGCTTCGGGCAATGCTTCATTTAATCTAAGCGACCACGCCG 2500
860 yTyrAspLeuThrGlyAspAsnAlaAlaLeuGluIleThrArgGlyAlaH 877
2501 TACAAAACGGCAGTCTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAGC 2550



```
    4225 GCTCAG.........GATTTCGGCAAAACCCGCAGTGCG...GAATGGGG 4262
         ::::::         ::::::|||||||||   |||:::   :::|||::
    1277 GlyArgThrGlyValGluLeuGlyLysThrPheSerGlyLysAspTrpSe 1293

    4263 CGTAAACGCC 4272
         :|||:::|||
    1293 rValThrAla 1296
```

seq_name: sp_bacteria:Q9LA54

seq_documentation_block:

```
ID   Q9LA54      PRELIMINARY;      PRT;  1335 AA.
AC   Q9LA54;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   EAAC.
GN   EAAC.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ECOR-9;
RX   MEDLINE=20187497; PubMed=10722621;
RA   Sandt C.H., Hill C.W.;
RT   "Four different genes responsible for nonimmune immunoglobulin-binding
RT   activities within a single strain of Escherichia coli.";
RL   Infect. Immun. 68:2205-2214(2000).
DR   EMBL; AF151674; AAF63038.1; -.
DR   MEROPS; S06.UPW; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
SQ   SEQUENCE   1335 AA;  141672 MW;  DC3EBB4E6AE4C969 CRC64;
```

alignment_scores:

```
            Quality: 1056.50              Length:   1504
              Ratio:  1.314                 Gaps:     65
 Percent Similarity:  53.457   Percent Identity: 27.261
```

alignment_block:

US-09-303-518D-649 x Q9LA54 ..

Align seg 1/1 to: Q9LA54 from: 1 to: 1335

```
       8 CAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGC 57
         :::   ::::::|||:::   |||   ::::::   |||         ||
      20 GluThrAlaGlyGlyArgLysLysSerSerGlyIlePro......GlyAl 34

      58 CGCATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCAT 107
         |:::      ||||||||||||           |||      :::::::
      34 aGlyMetLeuSerArgLeuLeuLeuAlaSerGlyAlaValAlaGlyValL 51

     108 TCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGGC....ATCAACTAC 153
         ::|||      |||:::      ::::::::::::      :::   :::
      51 euPheSerTyrProSerLeuAlaSerValValGlyAsnThrLeuProTrp 67

     154 CAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGC 203
         |||   |||||||||||||||||||||||||||   |||   :::||||||
      68 GlnThrTyrArgAspPheAlaGluAsnLysGlyAlaPheHisAlaGlyAl 84

     204 GAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAA 253
         |   :::|||   :::|||::::::::::|||      ||||||      :
      84 aThrAsnIleProLeuTyrGlyArgAsnGly...AlaValGlyGlyArgL 100

     254 TGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGT...AACGGC 300
         ::   ||||||||||||:::|||||||||||||||||::::::      |||
     100 euAspLysAlaProMetMetAspPheSerValValAspGlnIleLeuGly 116

     301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGG 350
         ||||||:::|||::::::   ||||||::::::::::|||   ||||||||
     117 ValAlaThrLeuIleSerProGlnTyrValAlaGlyValLysHisAsnGl 133

     351 CGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAAC 400
         |:::|||||:::|||   ||||||          :::
     133 ySerTyrAsnThrValArgPheGlyTyrAlaAspAspThr..... 146

     401 ATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCAGGGACT 450
                |||||||::::::|||   |||||||::
     147 .......ThrTyrArgLeuValAspArgAsnGlu..... 155

     451 AAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAATT 500
                      ::      |||:::|||   ||||||||||:::|||::
     156 ...........HisTrpArgAspPheHisThrProArgLeuAsnLysLe 168

     501 TGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGG. 549
         :||||||::::::   ||||||::::::::|||:::         |||:::
     168 uValThrGluValAlaProValSerValThrAspAlaGlyThrGlyLysG 185

     550 ..AAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCA 597
           |||   ::::::::::::||||||         |||:::|||:::
     185 lyValTyrGlnAsnArgSerArgTyrProValPheTyrArgMetGlySer 201

     598 GGCAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAG 647
         |||   ||||||     :::   :::                    ::
     202 GlyThrGlnTyrThrGlyAlaAlaSerGly.................Al 212

     648 TTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCT 697
         :      :::|||||||:::||||||::|||   :::||||||:::::
     212 aLeuThrArgIleAlaGlyAlaTyrAlaTrpLysThrGlyGlyThrValG 229

     698 TTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAA 747
           :::         |||      :::   |||:::   :::
     229 lySerProLeuIleSerAspTrpSerLeuValSer...AsnProGlyTyr 244

     748 ATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGA 797
         :::   :::|||      |||   |||   :::   |||:::   |||||
     245 LeuTyrGlnSerValAsnGlyProLeuAlaSerTyrGlyThrProGlyAs 261

     798 CAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAA 847
         ||||||||||||||:::|||   :::||||||   |||:::::|||::::
     261 pSerGlySerProLeuPheAlaTrpAspAlaValLysLysGlnTrpValL 278

     848 TTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGC 897
         ::   :::||||||          |||   |||   |||:::::::::
     278 euValAlaValLeu.........AsnGlyTyrAlaGlyGluLysGly... 290

     898 TTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGA 947
                   :::   :::||||||          |||   ||||||||
     291 ............LysThrAsnTrpPhe...ThrValIleProAlaGlyAs 302

     948 TACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTA 997
         |:::::::::                                 :::::::
     302 pValAsnAsn................................ThrIleL 308

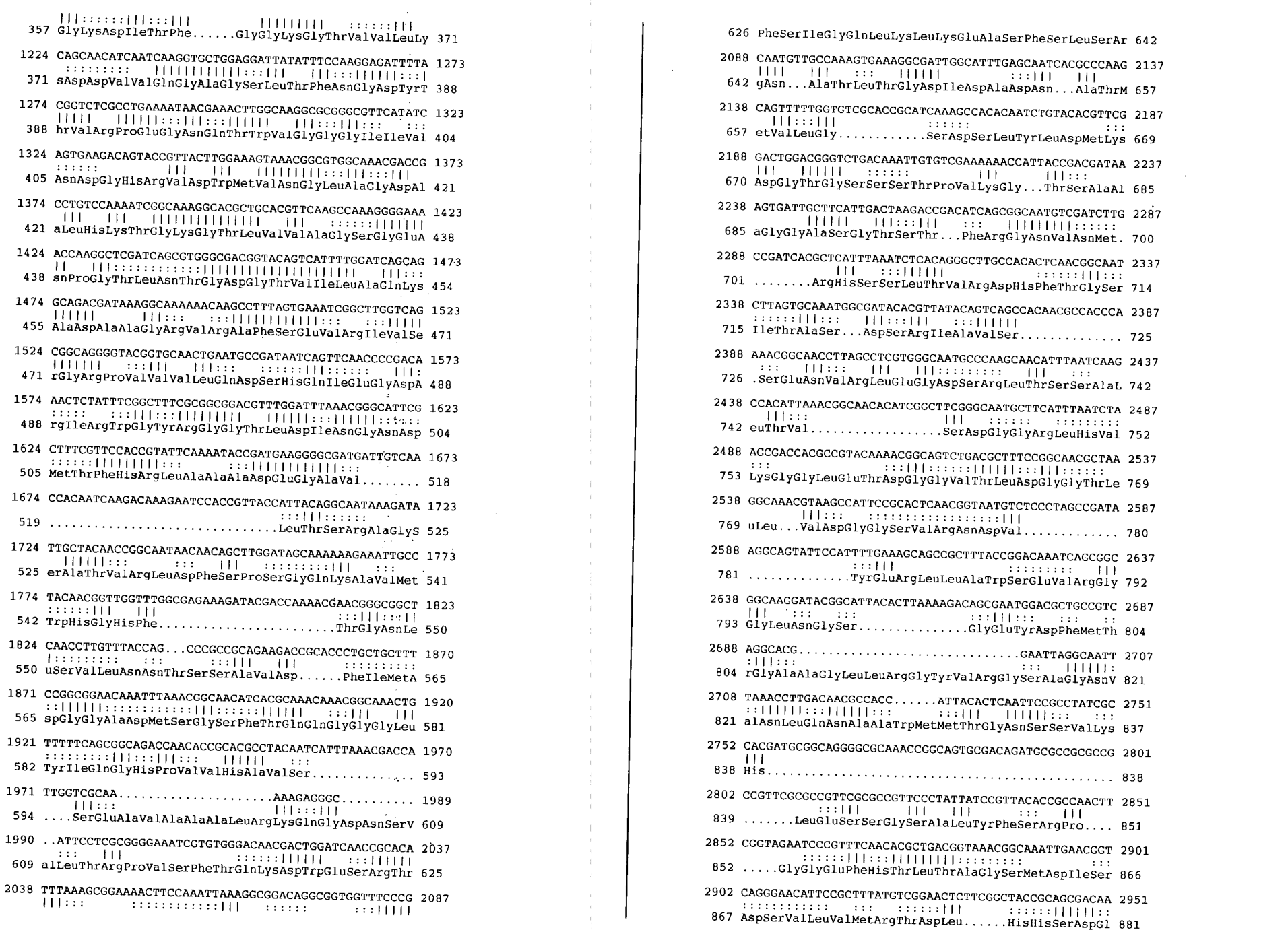
     998 ACGACGATAATAATGGCACA.....................GGAAAAATC 1026
         ::::::|||::::::||||||                    |||   |||
     308 ysGlnAspSerSerGlyThrValValProAlaValAlaGlyGlyAspIle 324

    1027 AATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAAC 1076
                :::::::::   :::|||      :::
     325 ValTrpAsnTyrSerLysGlySerGlyGluGly.................. 335

    1077 CGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTT 1126
                          :::||||||:::   ::::::
     336 ..................ThrLeuSerGlnAspGlyLys.......... 342

    1127 ATCATGCTGCAGGTGGTGTCAACAGTTATCGA...CCCAGACTGAATAAT 1173
                :::      :::|||::::::::|||      |||||||:::
     343 ........ValTrpLysMetAsnGlyPheArgGlyGlySerLeuAsnAsp 356

    1174 GGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTAC 1223
```



```
2952 ATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACA 3001
     :|||::::::::::|||||   :::|||                 |||||
 881 nLeuArgValThrGluSerLeuArgGlyLys.............AsnA 893

3002 ATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGA 3051
     ||                            |||   :::|||:::
 893 sn..........................LeuLeuLeuValAspPhe 899

3052 AAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGA 3101
       ::::::::      ::::::   |||||||:::   |||   :::
 900 ThrGluArgSerAspGlyGlnLysAlaLeuAsnIleProLeuValThrAl 916

3102 ACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCG 3151
        :::   :::||||||      :::::::::      :::|||
 916 aProAlaGlyThrGlyAlaAspValPheSerValLysThrArgAspThrG 933

3152 AGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTC 3201
       |||         :::|||||||   :::   |||
 933 lyPheSerHisIleThrProValValArgAlaGlu.............. 944

3202 GGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCT 3251
                          |||:::         :::|||   :::||
 945 ..................GlnGlyThrGlyGlyThrAlaTrpGlnLe 954

3252 TGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCG 3301
     |:::                          |||   :::   |||:::|
 954 uAsn.........................ValValGlnProGluThrV 962

3302 TTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAG 3351
     ||||||||||
 962 alAlaGluPro..................................... 965

3352 GCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGC 3401
     :::   :::|||   :::|||             :::     :::::
 966 ValValAspGluValSerArg............ProSerLeuProValVa 978

3402 GAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCG 3451
     :   :::::::::|||:::|||         :::   |||
 978 lMetArgGlnAspAlaLysThrProAsnProValGluAla.......... 991

3452 CCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCC 3501
                                   :::::::|||   |||:::|||
 992 ..........................ValSerProValProSerPro 998

3502 CAACCGCAGCGCGACCTG......ATCAGCCGTTATGCCAATAGCGGTTT 3545
        |||      ::::::      :::                |||:::|||
 999 AlaProAlaAlaSerValThrAspValProThrGlyGluAsnThrGlyGl 1015

3546 GAGT...GAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACG 3592
       |||   |||:::                              :::|||||
1015 uSerGlyGluTyr..........................ArgAspG 1022

3593 AATTAGACCGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGC 3642
     ||      |||   :::            |||::::::|||   :::|||
1022 lu...ThrArgTrpLeuLeuThrGlyTyrArgSerThrValAsnValSer 1037

3643 GGCATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTAC.. 3690
     ::::::||||||:::                |||:::      |||   :::
1038 AlaValArgAspAlaGlyMetLeuMetSerMetGlyHisArgAsnPheIl 1054

3691 ....................CGCCAACAAACCGACCTGCGCCAAATCG 3718
                          :::::::::   |||||||||:::|||:
1054 eAsnGluValAsnAsnLeuAsnLysArgMetGlyAspLeuArgAspIleA 1071

3719 GTATGCAGAAAAACCTC.................GGCAGCGGGCGCGTC 3750
     ::   :::   :::                     |||:::|||   :::
1071 snGlyGluAlaGlyAlaTrpAlaArgIleMetSerGlyThrGlySerAla 1087

3751 GGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTCGAC........ 3792
     |||      ||||||   |||   |||:::
1088 GlyGlyGlyPheSerAspAsnTyrThrHisValGlnValGlyAlaAspLy 1104

3793 .............GACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCG 3829
                    ||||||:::          :::   ::::::
1104 sLysHisGluLeuAspGlyLeuAspLeuPheThrGlyValThrMetThrT 1121

3830 CCGTTTTCGGGCAATACGGCATCGACAGGTTCTAC............... 3864
       :::   ::::::   |||   |||   |||
1121 yrThrAspSerAsnAlaGlySerAspThrPheSerGlyLysThrLysSer 1137

3865 ......ATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCAGCCTTTC 3908
              |||:::   |||:::|||   |||:::||||||:::
1138 ValGlyAlaGlyLeuTyrAlaSerAlaLeuPheAspSerGlyAlaTyrIl 1154

3909 AGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTACGGCATTC 3958
       |||   ||||||               :::   :::|||:::      :
1154 eAspLeuIleGly..............LysTyrValHisHis......A 1164

3959 AGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACCGCACATC 4008
     ::   :::|||   |||   |||:::|||:::|||   :::
1164 spAsnGluTyrThrAlaThrPheAlaGlyLeuGlyThrLys......... 1177

4009 GGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACGAAAACGT 4058
                                  ||||||
1178 ..............................AspTyr.............. 1179

4059 CAATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTA 4108
                              :::   :::   :::   ||||||   :
1180 .....................GlySerHisSerTrpTyrAlaGlyAlaG 1189

4109 AGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTAT 4158
     :::::   |||   :::      ::::::          |||   |||
1189 luValGlyTyrArgTyrHisValThrGluAspAlaTrpIleGluProGln 1205

4159 TTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAACA..... 4203
          :::|||   |||      |||:::|||||||||      :::
1206 AlaGluLeuValTyr...GlyAlaValSerGlyLysGlnPheSerTrpLy 1221

4204 .......CGCGTCAATACCGCCGTATTGGCTCAGGATTTC.......... 4236
                :::|||   ::::::    :::||||||
1221 sAspGlnGlyMetAsnLeuThrMetLysAspLysAspPheAsnProLeuI 1238

4237 ......................GGCAAAACCCGCAGTGCG...GAATGG 4260
                            |||||||:::   |||:::   :::|||
1238 leGlyArgThrGlyValAspValGlyLysSerPheSerGlyLysAspTrp 1254

4261 GGCGTAAACGCC 4272
       |||:::|||
1255 LysValThrAla 1258
```

seq_name: sp_bacteria:Q9LA58

seq_documentation_block:

```
ID   Q9LA58      PRELIMINARY;      PRT;  1335 AA.
AC   Q9LA58;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   EAAA.
GN   EAAA.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ECOR-9;
RX   MEDLINE=20187497; PubMed=10722621;
RA   Sandt C.H., Hill C.W.;
RT   "Four different genes responsible for nonimmune immunoglobulin-binding
RT   activities within a single strain of Escherichia coli.";
RL   Infect. Immun. 68:2205-2214(2000).
DR   EMBL; AF151091; AAF63237.1; -.
DR   MEROPS; S06.UPW; -.
DR   InterPro; IPR000710; IgA_S6.
```

```
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
SQ   SEQUENCE   1335 AA;  141677 MW;  B7876C33536F7411 CRC64;


alignment_scores:
           Quality: 1054.50                Length:   1511;
             Ratio:  1.318                   Gaps:     62
 Percent Similarity:  52.945   Percent Identity: 27.399

alignment_block:
 US-09-303-518D-649 x Q9LA58   ..

 Align seg 1/1  to: Q9LA58  from: 1  to: 1335

       8 CAACCGACAAACGGACAACCGAAACACACCGCAAAGCCCCGAAAACCGGC 57
         :::   ::::::|||:::   |||   ::::::   |||       .  ||
      20 GluThrAlaGlyGlyArgLysLysSerSerGlyIlePro......GlyAl 34

      58 CGCATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCAT 107
         |:::      ||||||||||||              |||      :::::::
      34 aGlyMetLeuSerArgLeuLeuLeuAlaSerGlyAlaValAlaGlyValL 51

     108 TCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGGC....ATCAACTAC 153
         ::|||      |||:::      ::::::::::::      :::   :::
      51 euPheSerTyrProSerLeuAlaSerValValGlyAsnThrLeuProTrp 67

     154 CAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGC 203
         |||   ||||||||||||||||||||||||||||   |||   :::||||||
      68 GlnThrTyrArgAspPheAlaGluAsnLysGlyAlaPheHisAlaGlyAl 84

     204 GAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAA 253
         |   :::|||   :::|||:::::::::|||      ||||||      :
      84 aThrAsnIleProLeuTyrGlyArgAsnGly...AlaValGlyGlyArgL 100

     254 TGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGT...AACGGC 300
         ::   ||||||||||||:::||||||||||||||||:::::      |||
     100 euAspLysAlaProMetMetAspPheSerValValAspGlnIleLeuGly 116

     301 GTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGG 350
         |||||||:::|||:::::   ||||||::::::::|||   ||||||||
     117 ValAlaThrLeuIleSerProGlnTyrValAlaGlyValLysHisAsnGl 133

     351 CGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAAC 400
         |:::|||||||:::|||   ||||||          :::
     133 ySerTyrAsnThrValArgPheGlyTyrAlaAspAspThr..... 146

     401 ATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCAGGGACT 450
               ||||||:::::|||   ||||||:::
     147 .......ThrTyrArgLeuValAspArgAsnGlu..... 155

     451 AAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAATT 500
                    :::      |||:::|||   ||||||||||:::|||::
     156 ...........HisTrpArgAspPheHisThrProArgLeuAsnLysLe 168

     501 TGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGG. 549
         :||||||:::::   ||||||:::::|||:::         |||:::
     168 uValThrGluValAlaProValSerValThrAspAlaGlyThrGlyLysG 185

     550 ..AAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCA 597
             |||   ::::::::::::||||||          |||:::|||:::
     185 lyValTyrGlnAsnArgSerArgTyrProValPheTyrArgMetGlySer 201

     598 GGCAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAG 647
         |||   ||||||      :::   :::                    ::
     202 GlyThrGlnTyrThrGlyAlaAlaSerGly.................Al 212

     648 TTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCT 697
         :      :::|||||||:::||||||:::|||   :::||||||:::::
     212 aLeuThrArgIleAlaGlyAlaTyrAlaTrpLysThrGlyGlyThrValG 229

     698 TTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAA 747
           :::         |||      :::   |||:::   :::

     229 lySerProLeuIleSerAspTrpSerLeuValSer...AsnProGlyTyr 244

     748 ATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGA 797
         :::   :::|||      |||   |||   :::   |||:::   |||||
     245 LeuTyrGlnSerValAsnGlyProLeuAlaSerTyrGlyThrProGlyAs 261

     798 CAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAA 847
         ||||||||||||||:::|||   :::|||||||   |||:::::|||:::
     261 pSerGlySerProLeuPheAlaTrpAspAlaValLysLysGlnTrpValL 278

     848 TTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGGC 897
         ::   :::|||||||          |||   |||   |||:::::::::
     278 euValAlaValLeu.........AsnGlyTyrAlaGlyGluLysGly... 290

     898 TTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGA 947
                     :::   :::|||||||          |||   ||||||||
     291 ............LysThrAsnTrpPhe...ThrValIleProAlaGlyAs 302

     948 TACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTTA 997
         |:::::::::
     302 pValAsnAsn.................................ThrIleL 308

     998 ACGACGATAATAATGGCACA.....................GGAAAAATC 1026
         :::::|||:::::|||||                        |||   |||
     308 ysGlnAspSerSerGlyThrValValProAlaValAlaGlyGlyAspIle 324

    1027 AATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAAC 1076
               :::::::::  :::|||      :::
     325 ValTrpAsnTyrSerLysGlySerGlyGluGly................. 335

    1077 CGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTT 1126
                         :::||||||::   ::::::
     336 ..................ThrLeuSerGlnAspGlyLys..... 342

    1127 ATCATGCTGCAGGTGGTGTCAACAGTTATCGA...CCCAGACTGAATAAT 1173
               :::      :::|||:::::|||         |||||||:::
     343 ........ValTrpLysMetAsnGlyPheArgGlyGlySerLeuAsnAsp 356

    1174 GGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTAC 1223
         |||:::::::|||:::|||         |||||||||   ::::::|||
     357 GlyLysAspIleThrPhe......GlyGlyLysGlyThrValValLeuLy 371

    1224 CAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTA 1273
         :::::::::   ||||||||||||||:::|||   |||:::||||||:::|
     371 sAspAspValValGlnGlyAlaGlySerLeuThrPheAsnGlyAspTyrT 388

    1274 CGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATC 1323
         |||||   ||||||:::|||:::||||||   |||:::|||:::   :::
     388 hrValArgProGluGlyAsnGlnThrTrpValGlyGlyGlyIleIleVal 404

    1324 AGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCG 1373
         ::::::          |||   |||   |||||||||:::|||:::|||
     405 AsnAspGlyHisArgValAspTrpMetValAsnGlyLeuAlaGlyAspAl 421

    1374 CCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAA 1423
          |||   |||   ||||||||||||||||   |||   :::::|||||||
     421 aLeuHisLysThrGlyLysGlyThrLeuValValAlaGlySerGlyGluA 438

    1424 ACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAG 1473
         ||   |||:::::::::::::|||||||||||||||||||||||   |||:::
     438 snProGlyThrLeuAsnThrGlyAspGlyThrValIleLeuAlaGlnLys 454

    1474 GCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAG 1523
         ||||||      |||:::   :::|||||||||||||:::   :::||||||
     455 AlaAspAlaAlaGlyArgValArgAlaPheSerGluValArgIleValSe 471

    1524 CGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACA 1573
         |||||||   :::|||   |||:::   ::::::|||::::::   |||:
     471 rGlyArgProValValValLeuGlnAspSerHisGlnIleGluGlyAspA 488

    1574 AACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCG 1623
         :::::   :::|||:::|||||||||   |||||||:::||||||::::::
     488 rgIleArgTrpGlyTyrArgGlyGlyThrLeuAspIleAsnGlyAsnAsp 504
```



```
970 lSerArgProSerLeuProValValMetArg.................G 981

3378 GGCGGATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGC 3427
981 lnAspValLysThrPro.........AsnProValGluAlaValSer... 993

3428 CGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGG.GATTTGCC 3476
994 ......ProValProSerProAlaProAlaAlaSerValThrAspValPr 1008

3477 GCAACTGCAACCCCAACCGCAGCCCCAACCGCAGCGCGAC.......... 3516
1008 oThrGlyGluAsnThrGlyGluSerGlyGluTyrArgAspGluThrArgT 1025

3517 ..CTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACG 3564
1025 rpLeuLeuThrGlyTyr.............................. 1030

3565 CTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGA 3614

1030 ..... 1030

3615 AGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAACACT 3664
1031 .......ArgSerThrValAsnAlaSerAlaValArgAspAlaGlyMetL 1045

3665 ACCGTTCGCAAGATTTCCGCGCCTAC..................... 3690
1045 euMetSerMetGlyHisArgAsnPheIleAsnGluValAsnAsnLeuAsn 1061

3691 CGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTC..... 3735
1062 LysArgMetGlyAspLeuArgAspIleAsnGlyGluAlaGlyAlaTrpAl 1078

3736 .............GGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACC 3772
1078 aArgIleMetSerGlyThrGlySerAlaGlyGlyGlyPheSerAspAsnT 1095

3773 GGACCGAAAACACCTTCGAC.....................GACGGCATC 3801
1095 yrThrHisValGlnValGlyAlaAspLysLysHisGluLeuAspGlyLeu 1111

3802 GGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCAT 3851
1112 AspLeuPheThrGlyValThrMetThrTyrThrAspSerAsnAlaGlySe 1128

3852 CGACAGGTTCTAC.....................ATCGGCATCAGCGCGG 3880
1128 rAspThrPheSerGlyLysThrLysSerValGlyAlaGlyLeuTyrAlaS 1145

3881 GCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGGCAAAATC 3930
1145 erAlaLeuPheAspSerGlyAlaTyrIleAspLeuIleGly..... 1158

3931 CGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTT 3980
1159 ......LysTyrValHisHis......AspAsnGluTyrThrAlaThrPh 1171

3981 CGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCC 4030
1171 eAlaGlyLeuGlyThrLysAspTyrSerSerHisSerTrpTyrAlaGlyA 1188

4031 AAAAAGCGGATTACCGCTAC.......................... 4050
1188 laGluValGlyTyrArgTyrHisValThrGluAspAlaTrpIleGluPro 1204

4051 ......GAAAACGTCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTA 4094
1205 GlnAlaGluLeuValTyrGlyAlaValSerGlyLysGlnPheSerTrpLy 1221

4095 CCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTT 4144

1221 sAspGlnGly..... 1224

4145 CCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAA 4194
1225 .............MetSerLeuSerMetLysAspLysAspTyrAsnPro 1236

4195 GTCCGAACACGCGTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAAC 4244
1237 LeuIleGlyArgThrGlyVal.............AspValGlyLysSe 1248

4245 CCGCAGTGCG...GAATGGGGCGTAAACGCC 4272
1248 rPheSerGlyLysAspTrpLysValThrAla 1258
```

seq_name: sp_bacteria:Q54165

seq_documentation_block:

```
ID   Q54165      PRELIMINARY;      PRT;  1366 AA.
AC   Q54165;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SEPA PROTEIN PRECURSOR.
GN   SEPA.
OS   Shigella flexneri.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Shigella.
OX   NCBI_TaxID=623;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=M90T;
RX   MEDLINE=96020667; PubMed=7476198;
RA   Benjelloun-Touimi Z., Sansonetti P.J., Parsot C.;
RT   "SepA, the major extracellular protein of Shigella flexneri:
RT   autonomous secretion and involvement in tissue invasion.";
RL   Mol. Microbiol. 17:123-135(1995).
CC   -!- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S1; ALSO KNOWN AS THE
CC       TRYPSIN FAMILY.
DR   EMBL; Z48219; CAA88252.1; -.
DR   MEROPS; S06.UPW; -.
DR   InterPro; IPR001314; Chymotrypsin.
DR   InterPro; IPR000710; IgA_S6.
DR   InterPro; IPR001254; Trypsin.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00722; CHYMOTRYPSIN.
DR   PROSITE; PS50240; TRYPSIN_DOM; 1.
KW   Hydrolase; Serine protease; Signal.
FT   SIGNAL       1     56       POTENTIAL.
FT   CHAIN       57   1089       MATURE SEPA.
SQ   SEQUENCE   1366 AA;  146379 MW;  85E8E05F3EF8A7D8 CRC64;
```

alignment_scores:

```
            Quality: 1001.50              Length:   1580
              Ratio:  1.253                 Gaps:     63
Percent Similarity:  50.570   Percent Identity: 25.886
```

alignment_block:

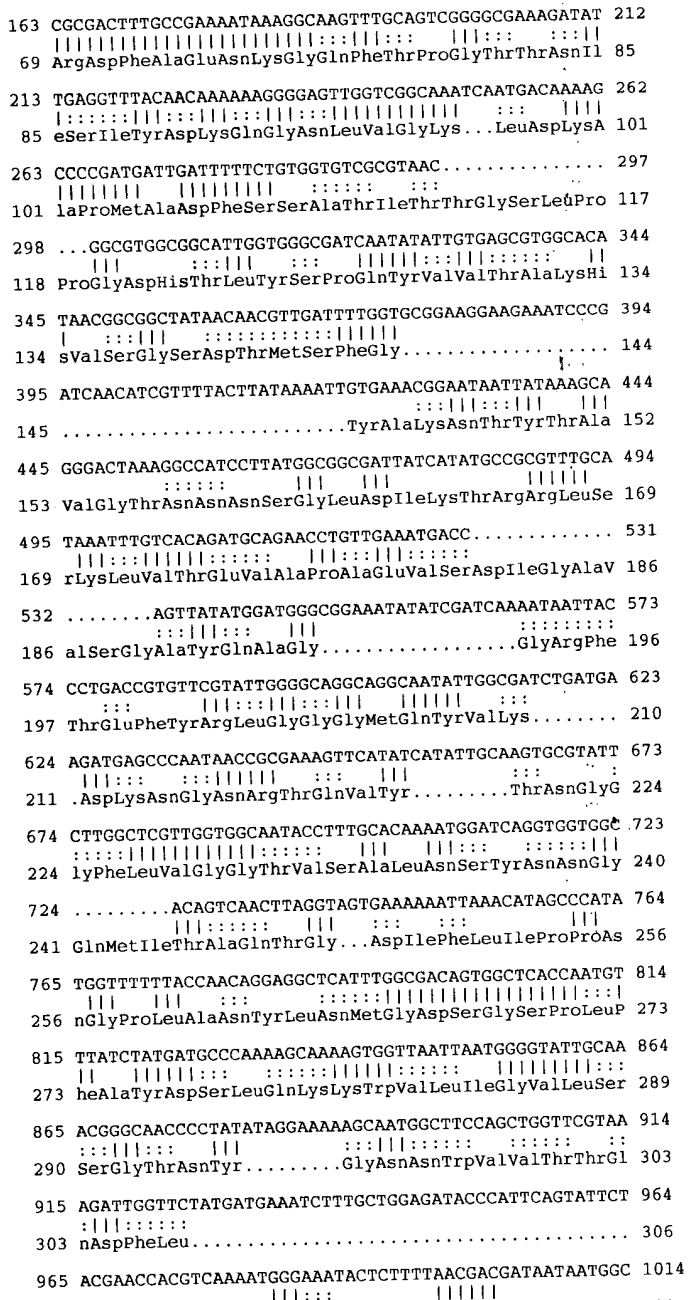
US-09-303-518D-649 x Q54165 ..

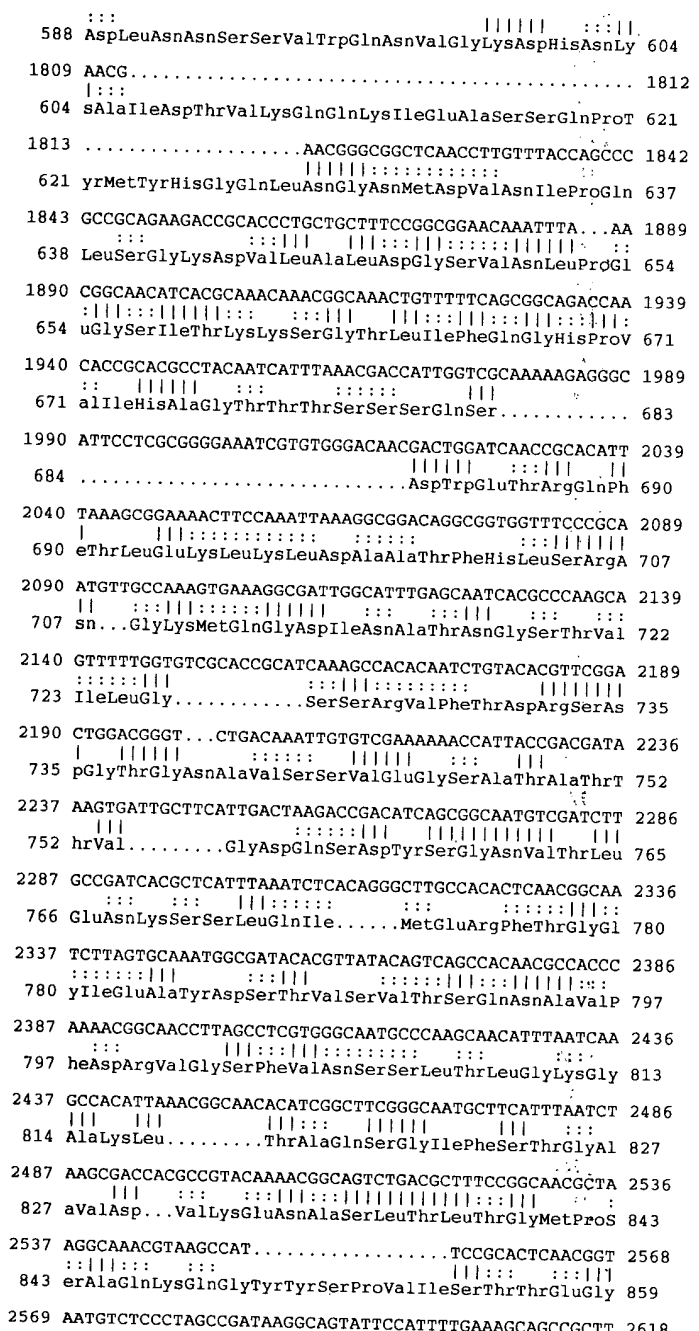
Align seg 1/1 to: Q54165 from: 1 to: 1366

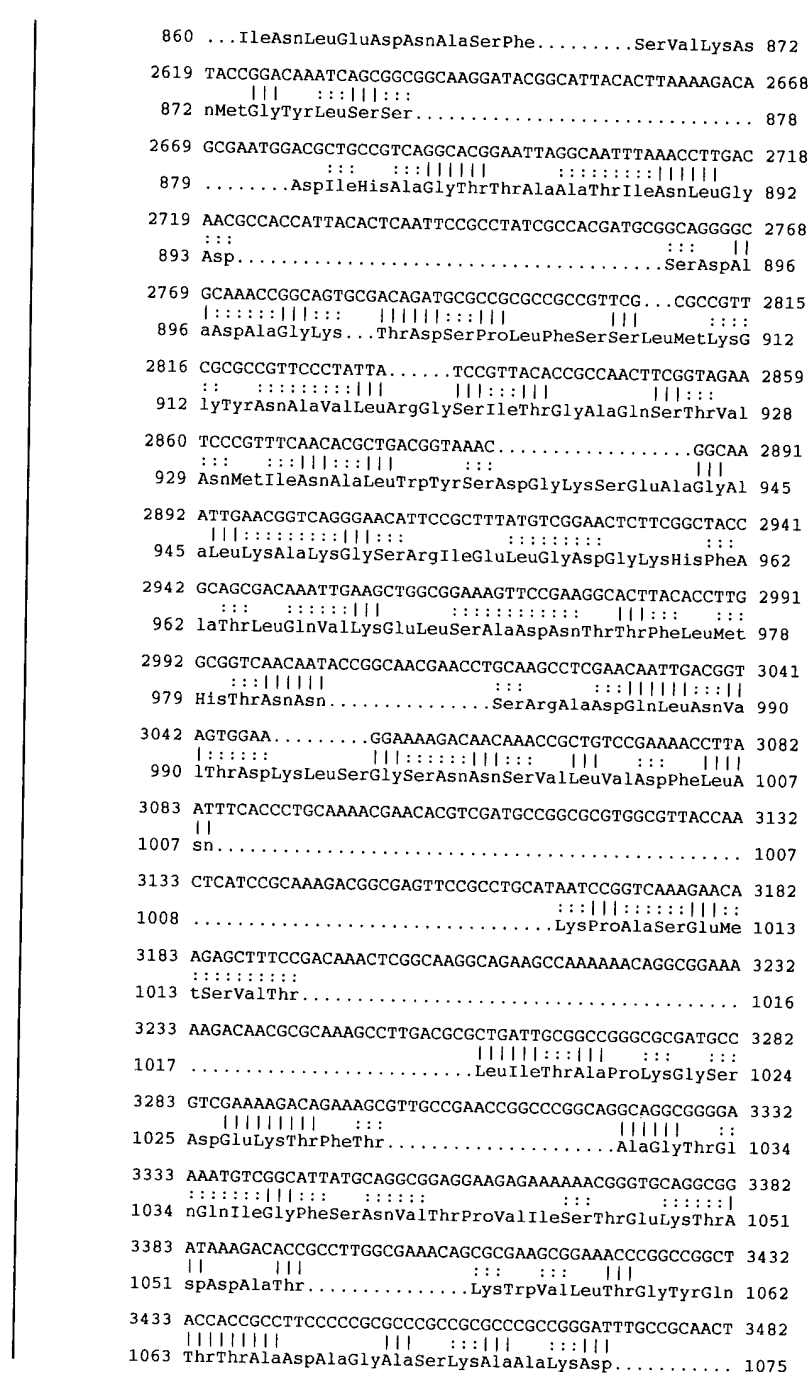
```
22 ACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGCATCCGCTTCTC 71
27 ThrCysLysSerHisArg.....................ArgLeuSe 35

72 GCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCCT 121
35 rArgArgValIleLeuThrSerValAlaAlaLeuSerLeuSerSerAlaT 52

122 GG.........GCGGGACACACTTATTTCGGCATCAACTACCAATACTAT 162
52 rpProAlaLeuSerAlaThrValSerAlaGluIleProTyrGlnIlePhe 68
```


163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATAT 212
69 ArgAspPheAlaGluAsnLysGlyGlnPheThrProGlyThrThrAsnIl 85
213 TGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAG 262
85 eSerIleTyrAspLysGlnGlyAsnLeuValGlyLys...LeuAspLysA 101
263 CCCCGATGATTGATTTTTCTGTGGTGTCGCGTAAC..... 297
101 laProMetAlaAspPheSerSerAlaThrIleThrThrGlySerLeuPro 117
298 ...GGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACA 344
118 ProGlyAspHisThrLeuTyrSerProGlnTyrValValThrAlaLysHi 134
345 TAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCG 394
134 sValSerGlySerAspThrMetSerPheGly..... 144
395 ATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCA 444
145 .....TyrAlaLysAsnThrTyrThrAla 152
445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
153 ValGlyThrAsnAsnAsnSerGlyLeuAspIleLysThrArgArgLeuSe 169
495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACC..... 531
169 rLysLeuValThrGluValAlaProAlaGluValSerAspIleGlyAlaV 186
532 .....AGTTATATGGATGGGCGGAAATATATCGATCAAAATAATTAC 573
186 alSerGlyAlaTyrGlnAlaGly.....GlyArgPhe 196
574 CCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGA 623
197 ThrGluPheTyrArgLeuGlyGlyGlyMetGlnTyrValLys..... 210
624 AGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATT 673
211 .AspLysAsnGlyAsnArgThrGlnValTyr.....ThrAsnGlyG 224
674 CTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGC 723
224 lyPheLeuValGlyGlyThrValSerAlaLeuAsnSerTyrAsnAsnGly 240
724 .....ACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATA 764
241 GlnMetIleThrAlaGlnThrGly...AspIlePheLeuIleProProAs 256
765 TGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGT 814
256 nGlyProLeuAlaAsnTyrLeuAsnMetGlyAspSerGlySerProLeuP 273
815 TTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAA 864
273 heAlaTyrAspSerLeuGlnLysLysTrpValLeuIleGlyValLeuSer 289
865 ACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAA 914
290 SerGlyThrAsnTyr.....GlyAsnAsnTrpValValThrThrGl 303
915 AGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCT 964
303 nAspPheLeu..... 306
965 ACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGC 1014
307 .....GlyGlnGlnProGlnAsnAsp..... 313
1015 ACAGGAAAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATT 1064
313 313
1065 AAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAA 1114
314 .....PheAspLysThrIleAlaTyrThrSerG 323
1115 GAGAACCTGTT.....TATCATGCTGCAGGTGGTGTCAAC... 1149
323 lyGluGlyValLeuGlnTrpLysTyrAspAlaAlaAsnGlyThrGlyThr 339
1150 .....AGTTATCGACCCAGA..... 1164
340 LeuThrGlnGlyAsnThrThrTrpGlyTyrAlaTrpLysGluArgLysLe 356
1165 ....CTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCG 1210
356 uIleLeuAsnAlaGlyLysAsnLeuLeuPheThrGly...AsnAsnGlyG 372
1211 AATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTC 1260
372 luValValLeuGlnAsnSerValAsnGlnGlyAlaGlyTyrLeuGlnPhe 388
1261 CAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGC 1310
389 AlaGlyAspTyrArgValSerAlaLeuAsnGlyGlnThrTrpMetGlyGl 405
1311 GGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCG 1360
405 yGlyIleIleThrAspLysGlyThrHisValLeuTrpGlnValAsnGlyV 422
1361 TGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAA 1410
422 alAlaGlyAspAsnLeuHisLysThrGlyGluGlyThrLeuThrValAsn 438
1411 GCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCAT 1460
439 GlyThrGlyValAsnAlaGlyGlyLeuLysValGlyAspGlyThrValIl 455
1461 TTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAA 1510
455 eLeuAsnGlnGlnAlaAspArgAspGlyLysValGlnAlaPheSerSerV 472
1511 TCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAG 1560
472 alGlyIleAlaSerGlyArgProThrValValLeuSerAspSerGlnGln 488
1561 TTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTT 1610
489 ValAsnProAspAsnIleSerTrpGlyTyrArgGlyGlyArgLeuGluLe 505
1611 AAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGG 1660
505 uAsnGlyAsnAsnLeuThrPheThrArgLeuGlnAlaAlaAspTyrGlyA 522
1661 CGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATT... 1707
522 laIleIleThrAsn...AsnSerGluLysLysSerThrValThrLeuAsp 537
1708 .....ACAGGCAATAAAGATATTGCTACAACCGGCAATAACAACAG 1748
538 LeuGlnThrLeuLysAlaSerAspIleAsnValProValAsnThrValSe 554
1749 CTTG.....GATAGCAAA. 1761
554 rIlePheGlyGlyArgGlyAlaProGlyAspLeuTyrTyrAspSerSerT 571
1762 .....AAAGAAATTGCCTACAACGGTTGGTTTGGC 1791
571 hrLysGlnTyrPheIleLeuLysAlaSerSerTyrSerProPhePheSer 587
1792 GAG.....AAAGATACGACCAA 1808


588 AspLeuAsnAsnSerSerValTrpGlnAsnValGlyLysAspHisAsnLy 604
1809 AACG 1812
604 sAlaIleAspThrValLysGlnGlnLysIleGluAlaSerSerGlnProT 621
1813 AACGGGCGGCTCAACCTTGTTTACCAGCCC 1842
621 yrMetTyrHisGlyGlnLeuAsnGlyAsnMetAspValAsnIleProGln 637
1843 GCCGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTA...AA 1889
638 LeuSerGlyLysAspValLeuAlaLeuAspGlySerValAsnLeuProGl 654
1890 CGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAA 1939
654 uGlySerIleThrLysLysSerGlyThrLeuIlePheGlnGlyHisProV 671
1940 CACCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGC 1989
671 alIleHisAlaGlyThrThrThrSerSerSerGlnSer 683
1990 ATTCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATT 2039
684 AspTrpGluThrArgGlnPh 690
2040 TAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCA 2089
690 eThrLeuGluLysLeuLysLeuAspAlaAlaThrPheHisLeuSerArgA 707
2090 ATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCA 2139
707 sn...GlyLysMetGlnGlyAspIleAsnAlaThrAsnGlySerThrVal 722
2140 GTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGA 2189
723 IleLeuGly.....SerSerArgValPheThrAspArgSerAs 735
2190 CTGGACGGGT...CTGACAAATTGTGTCGAAAAAACCATTACCGACGATA 2236
735 pGlyThrGlyAsnAlaValSerServalGluGlySerAlaThrAlaThrT 752
2237 AAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTCGATCTT 2286
752 hrVal.....GlyAspGlnSerAspTyrSerGlyAsnValThrLeu 765
2287 GCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCAA 2336
766 GluAsnLysSerSerLeuGlnIle.....MetGluArgPheThrGlyGl 780
2337 TCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACCC 2386
780 yIleGluAlaTyrAspSerThrValSerValThrSerGlnAsnAlaValP 797
2387 AAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAA 2436
797 heAspArgValGlySerPheValAsnSerSerLeuThrLeuGlyLysGly 813
2437 GCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTTAATCT 2486
814 AlaLysLeu.....ThrAlaGlnSerGlyIlePheSerThrGlyAl 827
2487 AAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAACGCTA 2536
827 aValAsp...ValLysGluAsnAlaSerLeuThrLeuThrGlyMetProS 843
2537 AGGCAAACGTAAGCCAT.....TCCGCACTCAACGGT 2568
843 erAlaGlnLysGlnGlyTyrTyrSerProValIleSerThrThrGluGly 859
2569 AATGTCTCCCTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTT 2618


860 ...IleAsnLeuGluAspAsnAlaSerPhe.....SerValLysAs 872
2619 TACCGGACAAATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACA 2668
872 nMetGlyTyrLeuSerSer 878
2669 GCGAATGGACGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGAC 2718
879 .....AspIleHisAlaGlyThrThrAlaAlaThrIleAsnLeuGly 892
2719 AACGCCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGC 2768
893 Asp.....SerAspAl 896
2769 GCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCG...CGCCGTT 2815
896 aAspAlaGlyLys...ThrAspSerProLeuPheSerSerLeuMetLysG 912
2816 CGCGCCGTTCCCTATTA.....TCCGTTACACCGCCAACTTCGGTAGAA 2859
912 lyTyrAsnAlaValLeuArgGlySerIleThrGlyAlaGlnSerThrVal 928
2860 TCCCGTTTCAACACGCTGACGGTAAAC.....GGCAA 2891
929 AsnMetIleAsnAlaLeuTrpTyrSerAspGlyLysSerGluAlaGlyAl 945
2892 ATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACC 2941
945 aLeuLysAlaLysGlySerArgIleGluLeuGlyAspGlyLysHisPheA 962
2942 GCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTG 2991
962 laThrLeuGlnValLysGluLeuSerAlaAspAsnThrThrPheLeuMet 978
2992 GCGGTCAACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGT 3041
979 HisThrAsnAsn.....SerArgAlaAspGlnLeuAsnVa 990
3042 AGTGGAA.....GGAAAAGACAACAAACCGCTGTCCGAAAACCTTA 3082
990 lThrAspLysLeuSerGlySerAsnAsnSerValLeuValAspPheLeuA 1007
3083 ATTTCACCCTGCAAAACGAACACGTCGATGCCGGCGCGTGGCGTTACCAA 3132
1007 sn 1007
3133 CTCATCCGCAAAGACGGCGAGTTCCGCCTGCATAATCCGGTCAAAGAACA 3182
1008 .....LysProAlaSerGluMe 1013
3183 AGAGCTTTCCGACAAACTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAA 3232
1013 tSerValThr 1016
3233 AAGACAACGCGCAAAGCCTTGACGCGCTGATTGCGGCCGGGCGCGATGCC 3282
1017 .....LeuIleThrAlaProLysGlySer 1024
3283 GTCGAAAAGACAGAAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGA 3332
1025 AspGluLysThrPheThr.....AlaGlyThrGl 1034
3333 AAATGTCGGCATTATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGG 3382
1034 nGlnIleGlyPheSerAsnValThrProValIleSerThrGluLysThrA 1051
3383 ATAAAGACACCGCCTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCT 3432
1051 spAspAlaThr.....LysTrpValLeuThrGlyTyrGln 1062
3433 ACCACCGCCTTCCCCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACT 3482
1063 ThrThrAlaAspAlaGlyAlaSerLysAlaAlaLysAsp..... 1075

```
3483 GCAACCCCAACCGCAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTATG 3532
                                                     :::
1076 ..............................................PheM 1077

3533 CCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCC 3582
       ||||||   :::::::|||   :::   :::|||::::::
1077 etAlaSerGlyTyrLysSerPheLeuThrGluValAsnAsnLeuAsnLys 1093

3583 GTACAGGACGAATTAGACCGCGTATTTGCCGAAGACCGCCGCAAC..... 3627
       :::   :::|||              :::|||   ::::::
1094 ArgMetGlyAspLeu..............ArgAspThrGlnGlyAspAl 1105

3628 .GCCGTTTGGACAAGCGGCATCCGGGACACCAAACACTACCGTTCGCAAG 3676
      :::|||||:::      ::::::   |||            |||   |
1105 aGlyValTrpAlaArgIleMetAsnGlyThr.........GlySerAlaA 1119

3677 ATTTCCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAG 3726
     ||         |||   ::::::           ||||||||:::::::
1119 sp...GlyAspTyrSerAspAsnTyrThrHisValGlnIleGlyValAsp 1134

3727 AAA.....          .....AACCTCGGCAGCGGGCGCGTCGGCAT 3755
     :::                 :::|||   :::|||   :::
1135 ArgLysHisGluLeuAspGlyValAspLeuPheThrGlyAlaLeuLeuTh 1151

3756 CCTGTTTTCGCACAACCGGACCGAAAACACCTTC.........GACGACG 3796
           :::   |||   ::::::::::::|||         :::   :
1151 rTyrThrAspSerAsnAlaSerSerHisAlaPheSerGlyLysAsnLysS 1168

3797 GCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATAC 3846
     :::::|||:::
1168 erValGlyGly..................................... 1171

3847 GGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTTTTAGCAG 3896
                          |||:::   |||:::|||   |||::::||
1172 ....................GlyLeuTyrAlaSerAlaLeuPheAsnSe 1181

3897 CGGCAGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGC 3946
     ||||:::      |||   ||||||               :::   ||||
1181 rGlyAlaTyrPheAspLeuIleGly..............LysTyrLeuH 1193

3947 ATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATC 3996
     ||:::      |||      :::   |||:::|||::::::::|||
1193 isHisAspAsnGln......HisThrAlaAsnPheAlaSerLeuGlyThr 1207

3997 GAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCG 4046
     :::   :::   :::   :::   |||          ::::::   |||||
1208 LysAspTyrSerSerHisSerTrpTyrAlaGlyAlaGluValGlyTyrAr 1224

4047 CTAC..................................GAAAACGTCA 4060
     ||||                                ||||   |||
1224 gTyrHisLeuThrLysGluSerTrpValGluProGlnIleGluLeuValT 1241

4061 ATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGCGCGGGCATTAAG 4110
           ::::::   |||   ||||||:::           |||:::
1241 yrGlySerValSerGlyLysAlaPheSerTrpGluAlaArgGlyMetAla 1257

4111 GCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCACGCCTTATTT 4160
               |||:::|||      ::::::      :::
1258 ......LeuSerMetLysAspLysAspTyrAsnProLeuIleGlyArgTh 1272

4161 GAGCCTGTCCTATACCGATGCCGCTTCGGGC.........AAAGTCCGAA 4201
      :::::::::         |||   ||||||         |||:::   :
1272 rGlyValAspValGlyArgAlaPheSerGlyAspAspTrpLysIleThrA 1289

4202 CACGC............................GTCAATACCGCCGTA 4221
     ::|||                          :::|||      :::
1289 laArgAlaGlyLeuGlyTyrGlnPheAspLeuLeuAlaAsnGlyGluThr 1305

4222 TTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGG........... 4260
     :::   ||||||   :::      :::      |||
1306 ValLeuGlnAspAlaSerGlyGluLysArgPheGluGlyGluLysAspSe 1322

4261 ..................GGCGTAAACGCCGAAATCAAA 4281
                         |||:::|||||||||||||||
1322 rArgMetLeuMetThrValGlyMetAsnAlaGluIleLys 1335
```

seq_name: sp_bacteria:Q99QC6

seq_documentation_block:

ID Q99QC6 PRELIMINARY; PRT; 1364 AA.
AC Q99QC6;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE SEPA, EXTRACELLULAR SERINE PROTEASE OF THE IGA1 PROTEASE FAMILY,
DE SECRETED BY A C-TERMINAL AUTOTRANSPORTER DOMAIN (SECRETED
DE PROTEASE).
GN SEPA.
OS Shigella flexneri.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Shigella.
OX NCBI_TaxID=623;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M90T;
RX MEDLINE=20566792; PubMed=11115111;
RA Buchrieser C., Glaser P., Rusniok C., Nedjari H., d'Hauteville H.,
RA Kunst F., Sansonetti P., Parsot C.;
RT "The virulence plasmid pWR100 and the repertoire of proteins secreted
RT by the type III secretion apparatus of Shigella flexneri.";
RL Mol. Microbiol. 38:760-771(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC PLASMID=VIRULENCE PWR501;
RA Venkatesan M.M., Goldberg M.B., Rose D.J., Grotbeck E.J., Burland V.,
RA Blattner F.R.;
RT "Complete DNA Sequence and Analysis of the Large Virulence Plasmid of
RT Shigella flexneri.";
RL Infect. Immun. 0:0-0(2001).
DR EMBL; AL391753; CAC05786.1; -.
DR EMBL; AF348706; AAK18385.1; -.
DR MEROPS; S06.UPW; -.
DR InterPro; IPR000710; IgA_S6.
DR InterPro; IPR001254; Trypsin.
DR Pfam; PF02395; IGA1; 1.
DR PRINTS; PR00921; IGASERPTASE.
DR SMART; SM00020; Tryp_SPc; 1.
DR PROSITE; PS50240; TRYPSIN_DOM; 1.
KW Hydrolase; Plasmid; Protease; Serine protease.
SQ SEQUENCE 1364 AA; 145937 MW; 71165A82BD098B9E CRC64;

alignment_scores:

Quality:	998.50	Length:	1585
Ratio:	1.254	Gaps:	63
Percent Similarity:	50.221	Percent Identity:	25.741

alignment_block:
US-09-303-518D-649 x Q99QC6 ..

Align seg 1/1 to: Q99QC6 from: 1 to: 1364

```
  22 ACAACCGAAACACACCGCAAAGCCCCGAAAACCGGCCGCATCCGCTTCTC 71
     |||   ::::::||||||                              |||:::||
  27 ThrCysLysSerHisArg.....................ArgLeuSe 35

  72 GCCTGCTTACTTAGCCATATGCCTGTCGTTCGGCATTCTTCCCCAAGCCT 121
     |         :::         ::::::          |||   :::||||
  35 rArgArgValIleLeuThrSerValAlaAlaLeuSerLeuSerSerAlaT 52

 122 GG.........GCGGGACACACTTATTTCGGCATCAACTACCAATACTAT 162
     ||         ::::::   :::         |||   ||||||   :::
  52 rpProAlaLeuSerAlaThrValSerAlaGluIleProTyrGlnIlePhe 68

 163 CGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGATAT 212
```

```
          ||||||||||||||||||||||:::|||:::   ||||:::   ::::||
 69 ArgAspPheAlaGluAsnLysGlyGlnPheThrProGlyThrThrAsnIl 85

213 TGAGGTTTACAACAAAAAGGGGAGTTGGTCGGCAAATCAATGACAAAAG 262
    |::::::|||:::|||:::|||:::|||||||||||   :::   ||||
 85 eSerIleTyrAspLysGlnGlyAsnLeuValGlyLys...LeuAspLysA 101

263 CCCCGATGATTGATTTTTCTGTGGTGTCGCGTAAC............... 297
    |||||||   |||||||||   ::::::   :::
101 laProMetAlaAspPheSerSerAlaThrIleThrThrGlySerLeuPro 117

298 ...GGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACA 344
       |||       :::|||   :::   ||||||:::|||::::::  .||
118 ProGlyAspHisThrLeuTyrSerProGlnTyrValValThrAlaLysHi 134

345 TAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCG 394
    |   :::|||   ::::::::::::||||||
134 sValSerGlySerAspThrMetSerPheGly................... 144

395 ATCAACATCGTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCA 444
                              :::|||:::|||   |||
145 .........................TyrAlaLysAsnThrTyrThrAla 152

445 GGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCA 494
             ::::::      |||   |||               ||||||
153 ValGlyThrAsnAsnAsnSerGlyLeuAspIleLysThrArgArgLeuSe 169

495 TAAATTTGTCACAGATGCAGAACCTGTTGAAATGACC............. 531
     |||:::|||||||::::::   |||:::|||::::::
169 rLysLeuValThrGluValAlaProAlaGluValSerAspIleGlyAlaV 186

532 ........AGTTATATGGATGGCGGAAATATATCGATCAAAATAATTAC 573
             :::|||:::   |||                  :::::::::
186 alSerGlyAlaTyrGlnAlaGly.....................GlyArgPhe 196

574 CCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGA 623
       :::      |||:::|||:::|||   ||||||   :::
197 ThrGluPheTyrArgLeuGlyGlyGlyMetGlnTyrValLys........ 210

624 AGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATT 673
     |||:::   :::||||||   :::   |||         :::   :  :
211 .AspLysAsnGlyAsnArgThrGlnValTyr.........ThrAsnGlyG 224

674 CTTGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGC 723
    ::::::||||||||||||||::::::   |||   |||:::   :::::|||
224 lyPheLeuValGlyGlyThrValSerAlaLeuAsnSerTyrAsnAsnGly 240

724 .........ACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATA 764
             |||::::::   |||:::
241 GlnMetIleThrAlaGlnThrGlyAspIle.................... 250

765 TGGTTTTTTACCAACAGGAGGC.................TCATTTGGCG 796
       |||   |||::::::|||                  ::::::||||
251 ....PheAsnProAlaAsnGlyProLeuAlaAsnTyrLeuAsnMetGlyA 266

797 ACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTA 846
    ||||||||||||:::|||   ||||||:::   ::::::||||||:::
266 spSerGlySerProLeuPheAlaTyrAspSerLeuGlnLysLysTrpVal 282

847 ATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGG 896
    :::   |||||||||:::::::|||   |||            :::|||::
283 LeuIleGlyValLeuSerSerGlyThrAsnTyr.........GlyAsnAs 296

897 CTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAG 946
    ::::   ::::::   :::|||::::::
296 nTrpValValThrThrGlnAspPheLeu...................... 305

947 ATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTT 996
                                        |||:::
306 ..................................GlyGlnGlnProGln 310

997 AACGACGATAATAATGGCACAGGAAAAATCAATGCCAAACATGAACACAA 1046
    ||||||
311 AsnAsp............................................ 312

1047 TTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTT 1096
                                              |||:::   :
313 ...................................PheAspLysT 316

1097 CTTTATCCGAGACAGCAAGAGAACCTGTT...........TATCATGCT 1134
     :::::::   |||:::   |||   |||             |||   |||
316 hrIleAlaTyrThrSerGlyGluGlyValLeuGlnTrpLysTyrAspAla 332

1135 GCAGGTGGTGTCAACAGTTATCGACCC...................... 1161
     |||:::|||:::::::::
333 AlaAsnGlyThrGlyThrLeuThrGlnGlyAsnThrThrTrpAspMetHi 349

1162 ................AGACTGAATAATGGAGAAAATATTTCCTTTATTG 1195
                      ||||||   |||:::|||:::   |||
349 sGlyLysLysGlyAsnAspLeuAsnAlaGlyLysAsnLeuLeuPheThrG 366

1196 ACGAAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTGCT 1245
           ::::::||||||::::::|||   :::::::::||||||||||||
366 ly...AsnAsnGlyGluValValLeuGlnAsnSerValAsnGlnGlyAla 381

1246 GGAGGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGA 1295
     |||   |||   |||   ||||||:::   ||||||      |||::::::
382 GlyTyrLeuGlnPheAlaGlyAspTyrArgValSerAlaLeuAsnGlyGl 398

1296 AACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTT 1345
     :||||||:::|||:::|||:::      ::::::   :::   |||   |
398 nThrTrpMetGlyGlyGlyIleIleThrAspLysGlyThrHisValLeuT 415

1346 GGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGC 1395
     ||:::||||||||||||||||:::|||:::|||   |||   |||:::|||
415 rpGlnValAsnGlyValAlaGlyAspAsnLeuHisLysThrGlyGluGly 431

1396 ACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGG 1445
     ||||||   |||::::::   |||   |||   |||:::::::::||||||
432 ThrLeuThrValAsnGlyThrGlyValAsnAlaGlyGlyLeuLysValGl 448

1446 CGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAAC 1495
     ||||||||||||||||||:::||||||||||||      ||||||   |
448 yAspGlyThrValIleLeuAsnGlnGlnAlaAspAlaAspGlyLysValG 465

1496 AAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTG 1545
     ||||||||||::::::|||::::::|||||||||   ||||||   |||
465 lnAlaPheSerSerValGlyIleAlaSerGlyArgProThrValValLeu 481

1546 AATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGG 1595
     :::   ::::::|||   ||||||||||::::::   :::|||:::||||||
482 SerAspSerGlnGlnValAsnProAspAsnIleSerTrpGlyTyrArgGl 498

1596 CGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAA 1645
     ||||||||||:::||||||||||::::::|||:::|||   |||:::|||
498 yGlyArgLeuGluLeuAsnGlyAsnAsnLeuThrPheThrArgLeuGlnA 515

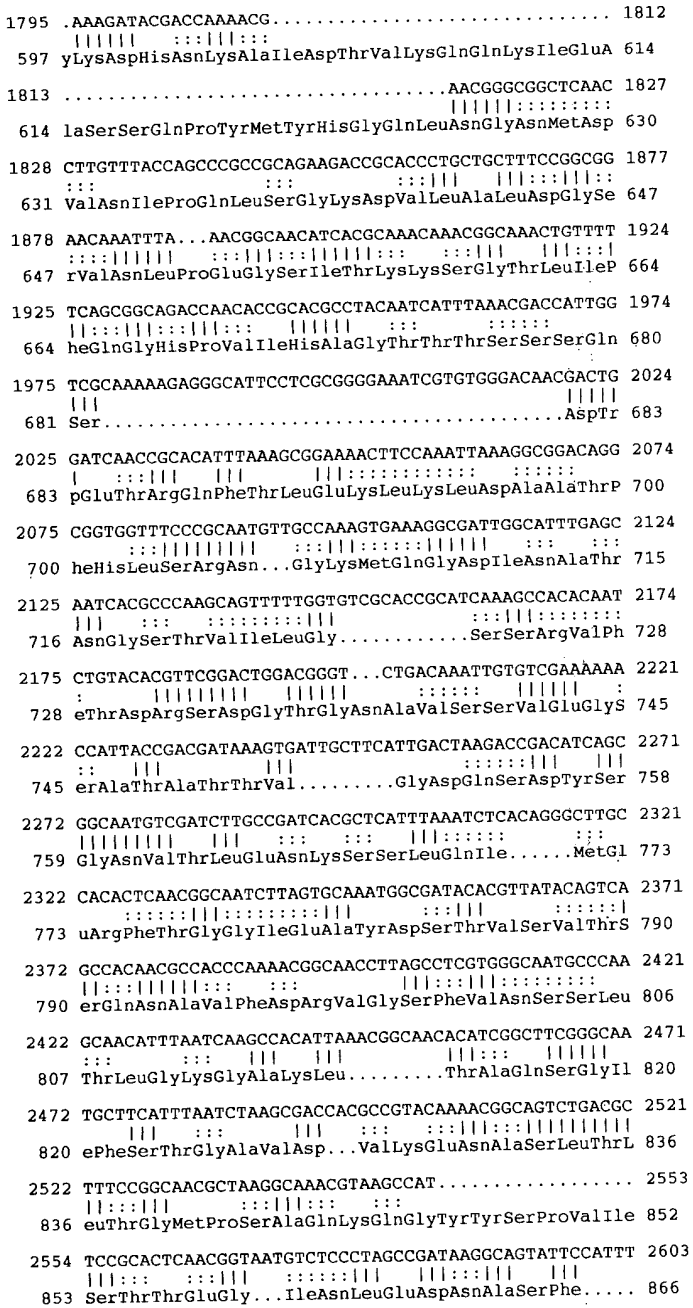
1646 ATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCC 1695
        :::|||   ||||||:::|||:::|||   |||::::::|||:::|||
515 laAlaAspTyrGlyAlaIleIleThrAsn...AsnSerGluLysLysSer 530

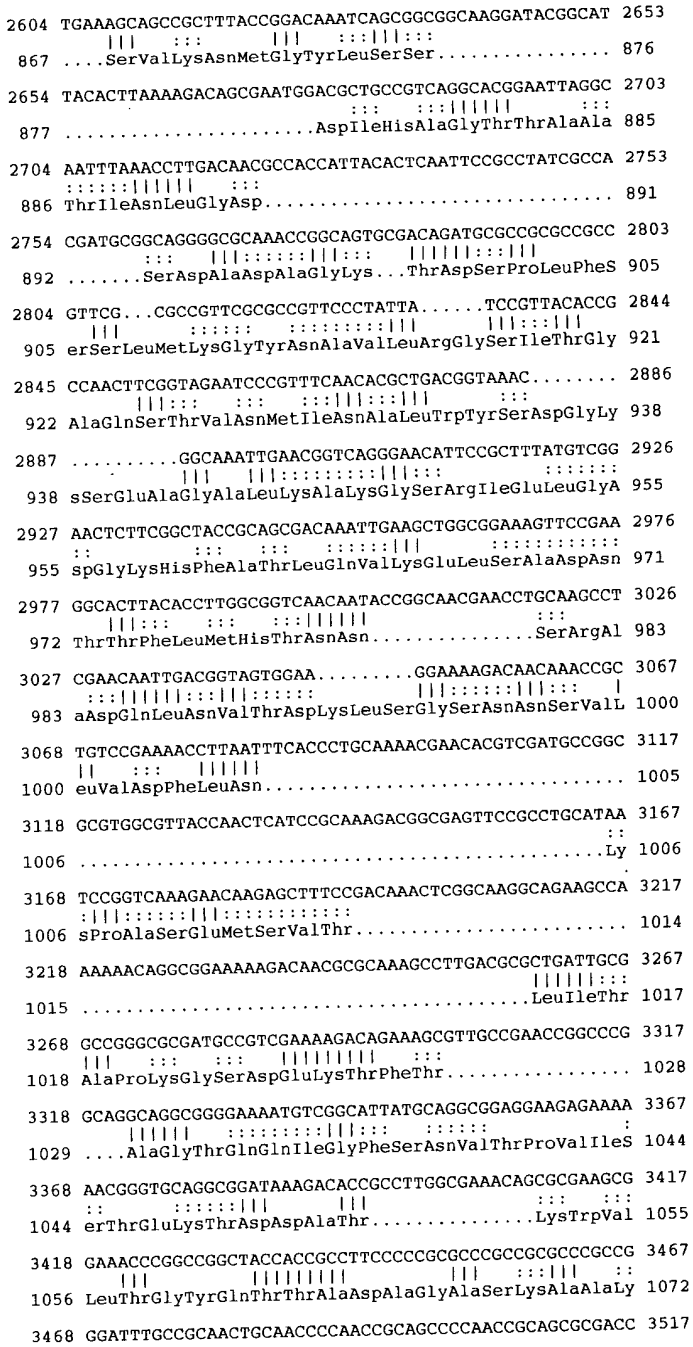
1696 ACCGTTACCATT...........ACAGGCAATAAAGATATTGCTACAAC 1733
     ||||||||||:::                     :::||||||   :::
531 ThrValThrLeuAspLeuGlnThrLeuLysAlaSerAspIleAsnValPr 547

1734 CGGCAATAACAACAGCTTG.............................. 1752
        |||:::   |||:::
547 oValAsnThrValSerIlePheGlyGlyArgGlyAlaProGlyAspLeuT 564

1753 .....GATAGCAAA....................AAAGAAATTGCCTAC 1776
        ||||||:::                   |||      :::|||
564 yrTyrAspSerSerThrLysGlnTyrPheIleLeuLysAlaSerSerTyr 580

1777 AACGGTTGGTTTGGCGAG............................... 1794
     :::   :::|||:::::
581 SerProPhePheSerAspLeuAsnAsnSerSerValTrpGlnAsnValGl 597
```





```
      :|||
1072 sAsp..................................................... 1073

3518 TGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCCACGCTC 3567
                :::      ||||||   :::::::|||   :::   :::
1074 ..........PheMetAlaSerGlyTyrLysSerPheLeuThrGluVal 1086

3568 AACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGCCGAAGA 3617
     |||::::::          :::   :::|||              :::||
1087 AsnAsnLeuAsnLysArgMetGlyAspLeu..............ArgAs 1098

3618 CCGCCGCAAC......GCCGTTTGGACAAGCGGCATCCGGGACACCAAAC 3661
     |   ::::::      :::|||||||::      ::::::   |||
1098 pThrGlnGlyAspAlaGlyValTrpAlaArgIleMetAsnGlyThr.... 1113

3662 ACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGC 3711
               |||  |||          |||   ::::::
1114 .....GlySerAlaAsp...GlyAspTyrSerAspAsnTyrThrHisVal 1127

3712 CAAATCGGTATGCAGAAA.....................AACCTCGGCAG 3740
     |||||||||:::::::::                     :::|||   ::
1128 GlnIleGlyValAspArgLysHisGluLeuAspGlyValAspLeuPheTh 1144

3741 CGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTC. 3789
     :|||  :::             :::  |||  ::::::::::::|||
1144 rGlyAlaLeuLeuThrTyrThrAspSerAsnAlaSerSerHisAlaPheS 1161

3790 ........GACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCC 3831
             :::   ::::::|||:::
1161 erGlyLysAsnLysSerValGlyGly..................... 1169

3832 GTTTTCGGGCAATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGG 3881
                                        |||:::   |||::
1170 ................................GlyLeuTyrAlaSe 1174

3882 CGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGGCAAAATCC 3931
     :|||  |||:::||||||:::     |||  ||||||
1174 rAlaLeuPheAsnSerGlyAlaTyrPheAspLeuIleGly.......... 1187

3932 GCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTC 3981
        :::  ||||||:::      |||      :::   |||:::|||
1188 .....LysTyrLeuHisHisAspAsnGln......HisThrAlaAsnPhe 1200

3982 GGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCA 4031
     ::::::::::|||  :::  :::  :::  :::  |||
1201 AlaSerLeuGlyThrLysAspTyrSerSerHisSerTrpTyrAlaGlyAl 1217

4032 AAAAGCGGATTACCGCTAC.............................. 4050
      ::::::  ||||||||
1217 aGluValGlyTyrArgTyrHisLeuThrLysGluSerTrpValGluProG 1234

4051 .....GAAAACGTCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTAC 4095
        |||  |||      ::::::  |||  ||||||:::
1234 lnIleGluLeuValTyrGlySerValSerGlyLysAlaPheSerTrpGlu 1250

4096 CGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTC 4145
          |||:::            |||:::|||      ::::::
1251 AspArgGlyMetAla......LeuSerMetLysAspLysAspTyrAsnPr 1265

4146 CATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGC.... 4191
     :::          :::::::::        |||   ||||||
1265 oLeuIleGlyArgThrGlyValAspValGlyArgAlaPheSerGlyAspA 1282

4192 .....AAAGTCCGAACACGC........................... 4206
        |||:::   :::|||
1282 spTrpLysIleThrAlaArgAlaGlyLeuGlyTyrGlnPheAspLeuLeu 1298

4207 GTCAATACCGCCGTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGA 4256
     :::|||      ::::::   ||||||  :::      :::      ||
1299 AlaAsnGlyGluThrValLeuGlnAspAlaSerGlyGluLysArgPheGl 1315

4257 ATGG.............................GGCGTAAACGCCGAAA 4276
     |                                 |||:::||||||||||
```

```
1315 uGlyGluLysAspSerArgMetLeuMetThrValGlyMetAsnAlaGluI 1332

4277 TCAAA 4281
     |||||
1332 leLys 1333
```

seq_name: sp_bacteria:Q93AF7

```
seq_documentation_block:
ID   Q93AF7      PRELIMINARY;      PRT;  1359 AA.
AC   Q93AF7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   AUTOTRANSPORTER.
GN   EPEA.
OS   Escherichia coli.
OG   Plasmid pO113.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Sloan J., Leyton D., Hartland E.L.;
RT   "Putative adhesin encoded by the large virulence plasmid, pO113, of
RT   enterohemorrhagic Escherichia coli O113:H21.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF422145; AAL18821.1; -.
KW   Plasmid.
SQ   SEQUENCE   1359 AA;  147339 MW;  A9C84FF1C662949C CRC64;
```

```
alignment_scores:
            Quality:  975.00              Length:   1534
              Ratio:  1.252                 Gaps:     63
 Percent Similarity:  50.782   Percent Identity: 25.489
```

```
alignment_block:
 US-09-303-518D-649 x Q93AF7    ..

 Align seg 1/1  to: Q93AF7  from: 1  to: 1359

      49 AAAACCGGCCGCATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTC 98
         ||||||||||||   ::::::      |||   :::|||:::   :::||
      29 LysThrGlyArg...LysLeuMetThrAlaSerValAlaLeuSerValSe 44

      99 GTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGGCATCA 148
         |::::::   ||||||   :::   |||:::   :::          |||
      44 rLeuSerAlaLeuPro...ValGluAlaSerThrValSerAlaGluIleP 60

     149 ACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTC 198
         ||||||   :::|||||||||||||||||||||||||||   |||:::
      60 roTyrGlnThrPheArgAspPheAlaGluAsnLysGlyValPheThrPro 76

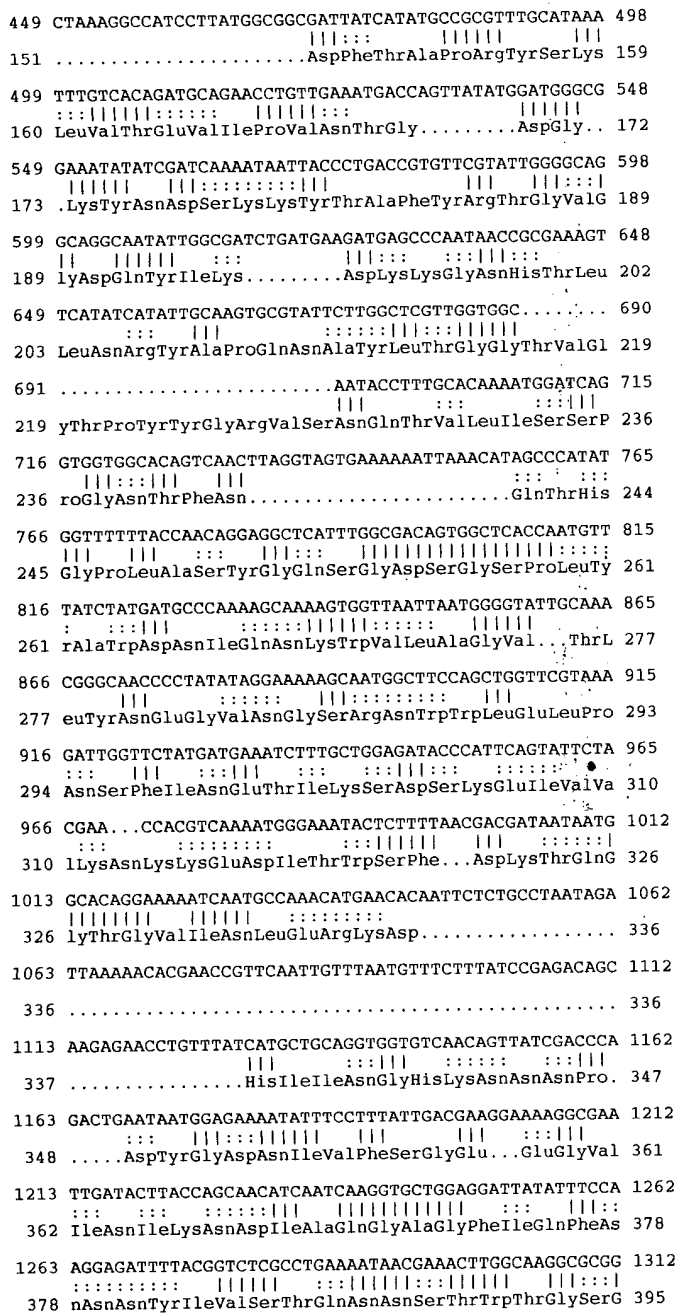
     199 GGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAA 248
         ||||||      |||||||:::   :::|||:::|||:::|||:::|||
      77 GlyAlaThrGlyIleGluIleLysAspLysAsnGlyAsnLeuThrGly.. 92

     249 ATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACG 298
             :::   ::::::||||||:::|||||||||   :::||||||:::|
      93 ....ValLeuAsnThrProMetLeuAspPheSerSerLeuSerArgAspG 108

     299 GCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAAC 348
         ||   ::::::|||:::         |||   ||||||:::   |||
     108 lyHisThrThrLeuIleHisProGlyTyrGlyValSerAlaLysHis... 123

     349 GGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCA 398
         |||      :::::::::      ::::::  |||  |||   :::
     124 GlyArgLeuGlnSerLeuLysThrAlaThrPheGlyTyrAsnAlaGlu.. 139

     399 ACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCAGGGA 448
                 ||||||||||||:::::::|||:::   :::
     140 .............TyrLysIleValAsnAsnAsnHisAlaSerTyr.... 150
```


449 CTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAA 498
151 .....AspPheThrAlaProArgTyrSerLys 159
499 TTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGGCG 548
160 LeuValThrGluValIleProValAsnThrGly.....AspGly.. 172
549 GAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAG 598
173 .LysTyrAsnAspSerLysLysTyrThrAlaPheTyrArgThrGlyValG 189
599 GCAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGT 648
189 lyAspGlnTyrIleLys.....AspLysLysGlyAsnHisThrLeu 202
649 TCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGC..... 690
203 LeuAsnArgTyrAlaProGlnAsnAlaTyrLeuThrGlyGlyThrValGl 219
691 .....AATACCTTTGCACAAAATGGATCAG 715
219 yThrProTyrTyrGlyArgValSerAsnGlnThrValLeuIleSerSerP 236
716 GTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATAT 765
236 roGlyAsnThrPheAsn.....GlnThrHis 244
766 GGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTT 815
245 GlyProLeuAlaSerTyrGlyGlnSerGlyAspSerGlySerProLeuTy 261
816 TATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAA 865
261 rAlaTrpAspAsnIleGlnAsnLysTrpValLeuAlaGlyVal...ThrL 277
866 CGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAA 915
277 euTyrAsnGluGlyValAsnGlySerArgAsnTrpTrpLeuGluLeuPro 293
916 GATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTA 965
294 AsnSerPheIleAsnGluThrIleLysSerAspSerLysGluIleValVa 310
966 CGAA...CCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATG 1012
310 lLysAsnLysLysGluAspIleThrTrpSerPhe...AspLysThrGlnG 326
1013 GCACAGGAAAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGA 1062
326 lyThrGlyValIleAsnLeuGluArgLysAsp..... 336
1063 TTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGC 1112
336 336
1113 AAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCA 1162
337 .....HisIleIleAsnGlyHisLysAsnAsnAsnPro. 347
1163 GACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAA 1212
348 .....AspTyrGlyAspAsnIleValPheSerGlyGlu...GluGlyVal 361
1213 TTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCA 1262
362 IleAsnIleLysAsnAspIleAlaGlnGlyAlaGlyPheIleGlnPheAs 378
1263 AGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCGCGG 1312
378 nAsnAsnTyrIleValSerThrGlnAsnAsnSerThrTrpThrGlySerG 395
1313 GCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTG 1362
395 lyIleIleValAsnAspGlySerGluValTyrTrpGlyIleAsnGlyVal 411
1363 GCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGC 1412
412 LysAspAspAsnLeuHisLysLeuGlySerGlyLysLeuIleIleAsnGl 428
1413 CAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTT 1462
428 yAsnGlyIleAsnThrGlyGlyLeuLysValGlyAspGlyThrValLeuL 445
1463 TGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATC 1512
445 euSerGlnHisLysAspAlaGluGlyArgValGlnAlaPheSerServal 461
1513 GGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTT 1562
462 AsnIleSerSerGlyArgProValValValLeuSerAspAspLysGlnIl 478
1563 CAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAA 1612
478 eAsnProAspAsnIleSerTrpGlyTyrArgGlyGlyThrLeuAspLeuA 495
1613 ACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCG 1662
495 snGlyAsnAsnIleThrPheThrArgIleGlnAlaSerAspTyrGlyAla 511
1663 ATGATTGTCAACCACAATCAAGACAAAGAA..... 1692
512 IleIleThrAsnAsnAsnAspLysLysSerIleThrAsnLeuAsnIleSe 528
1693 .....TCCACCGTTACCATTA 1708
528 rThrLeuLysAspThrAspIleAsnValSerValSerAspValHisValL 545
1709 CAGGCAATAAAGAT.....ATTGCTACAACCGGCAATAAC 1743
545 euGlyGlyLysGlyLysProGlyAspLeuPheTyrAspArgGlyThrGly 561
1744 AACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGA 1793
562 ArgTyrLeuIleLeuLysLysGlu...LysTyrSerProPhePheGlnAs 577
1794 GAAAGATACGACC..... 1806
577 pValAsnAsnThrSerValTrpGlnTyrValGlyLysAsnArgAsnGluA 594
1806 1806
594 laIleAlaValValLysSerGluLysIleLysAsnSerSerLysProTyr 610
1807 .....AAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGC 1844
611 LeuPheHisGlyGlnLeuAsnGlyAsnMetAspIleAsnIleGlnSerGl 627
1845 CGCAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTA...AACG 1891
627 yMetGlyValAsnThrPheAlaIleAspGlySerValAsnIleProAsnG 644
1892 GCAACATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAACA 1941
644 lyThrLeuSerAsnLysLeuGlySerLeuIlePheGlnGlyHisProVal 660
1942 CCGCACGCCTACAATCATTTAAACGACCATTGGTCGCAAAAAGAGGGCAT 1991
661 IleHisAla.....GlyVa 665
1992 TCCTCGCGGGGAAATCGTGTGGGACAACGACTGGATCAACCGCACATTTA 2041
665 l.....GlyValSerAlaGluGlnAsnAspTrpGluThrArgTyrPheT 680
2042 AAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAAT 2091

```
                ::::::::::::::::   |||::::::      :::||||||||||
 680 hrLeuAsnThrLeuAsnLeuAspGlySerAspPheHisLeuSerArgAsn 696

2092 GTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGT 2141
       |||   :::   |||:::   |||   :::|||   ::::::::::
 697 ...AlaValMetThrGlyAsnIleHisAlaThrAsnSerSerLysValTh 712

2142 TTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACACGTTCGGACT 2191
     ::::|||              :::|||   :::          |||:::|||
 712 rLeuGly............SerSerSerValTyrValAspArgAsnAspG 725

2192 GGACGGGT...CTGACAAATTGTGTCGAAAAAACCATTACCGACGATAAA 2238
       ||||||      :::|||   ::::::   |||   :::|||
 725 lyThrGlyGluSerValAsnSerThrSerAlaThrAlaSerAsp..... 739

2239 GTGATTGCTTCATTGACTAAGACCGACATCAGC.........GGCAATGT 2279
                  :::::::::   |||:::|||         |||:::::
 740 ............IleAsnGluLysAspLeuSerLysPheThrGlyAspIl 752

2280 CGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCA 2329
      :      ||||||:::::::::::::   :::::::::          ::::
 752 e...ThrAlaAspAsnSerAsnIleIleIleAsnAsn......LysPheG 766

2330 ACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTC...AGCCAC 2376
     ::|||:::::|||||||   :::   :::          :::   |||:::
 766 luGlyGlyValSerAlaValAsnLysSerAlaIleAspIleHisSerGln 782

2377 AACGCC.......................ACCCAAAACGGCAACCTTAG 2402
     :::|||                      ::::::|||:::::::|||::
 783 HisAlaValIleAsnArgTrpSerAspIleSerAspAsnSerLysLeuTh 799

2403 CCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTAAACGGCA 2452
     :|||      ||||||      |||   |||   :::   :::|||   :
 799 rLeuLysLysAsnAlaThrLeuThrValAsnThrGlyLeuValAsnLysG 816

2453 ACACATCGGCTTCGGGC...AATGCTTCATTTAATCTAAGCGACCACGCC 2499
     ::|||        |||   ||||||::::::|||||:::   ::::.
 816 lyThrIleGluIleGlyGluAsnAlaGluLeuAsnLeuGlnGlyTyrPro 832

2500 GTACAAAAC.........GGCAGTCTGACGCTTTCCGGCAACGCTAAGGC 2540
     :::   :::            ||||||         ||||||:::|||
 833 IleAlaAspLysPheIleProSerLeuHisAspLeuGlyAsnValLysMe 849

2541 AAACGTAAGCCATTCCGCACTCAAC...GGTAATGTCTCCCTAGCCGATA 2587
      ::::::|||:::::::::|||:::   ||||||   ::::::
 849 tThrAlaSerAsnAlaThrLeuThrAlaGlyAsnTyrAlaMet....... 863

2588 AGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGC 2637
                                 |||:::|||:::|||::::::
 864 ............................PheSerGlyGluIleThrAla 870

2638 GGCAAGGATACGGCATTACACTTAAAA...GACAGCGAATGGACGCTGCC 2684
           ||||||:::::::::::::      ||||||   :::
 871 AspAspAlaThrAlaValArgValAsnLeuGlySerGluThrSerThrLe 887

2685 GTCA............GGCACGGAATTAGGCAATTTAAACCTTGACAACG 2722
      |||             :::   ||||||   :::|||   :::|||:::
 887 uSerGluPheAsnProAsnProGluLeuThrAspLeuMetPheAspLysT 904

2723 CCACCATTACACTCAAT.........TCCGCCTATCGCCACGATGCGGCA 2763
       :::   :::   :::         ||||||   :::   ||||||:::
 904 yrAsnThrSerTrpThrGlyLysIleSerAlaLeuLysGlyAspAlaSer 920

2764 GGGGCGCAAACC...GGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCG 2810
       :::::::|||          |||                 ::::::
 921 MetValAsnThrValTrpArgMetThrGlyAspSerGlyLeuAsnThrLe 937

2811 CCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAAT 2860
     ::::::   :::|||||||                  |||   ::::
 937 uLysThrSerLysSerLeuThrValPhe............SerSerAspA 950

2861 CCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACA 2910
     :::::|||:::|||||||||||||||||
```

```
 950 snLysPheSerThrLeuThrValAsnAsp..................... 959

2911 TTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCT 2960

 959 ..................................................... 959

2961 GGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAATACCGGCA 3010
            :::|||::::::|||::::::|||    ::::::||||||:
 960 .LeuThrThrSerAspSerThrPheValLeuArgSerAspSerThrGlyS 976

3011 ACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAAC 3060
     :::::   :::   ::::::      :::   :::||||||||||:::|||
 976 erAspLysValValValLys......AsnLysLeuGluGlyLysAsnAsn 990

3061 AAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGA 3110
     :::   |||                                    |||||
 991 AsnLeuLeu.................................ValAs 995

3111 TGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCC 3160
     |                    :::   :::|||||||:::
 995 p.....................TyrValAlaAsnAspGlyLys..... 1002

3161 TGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCA 3210
       ||||||   ::::::   :::   :::|||          |||   |||
1003 ..HisAsnSerLeuAsnLeuGluLeuValSerAlaProLysGlyThrAla 1018

3211 GAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCT 3260
                  |||:::      |||:::|||:::   ::::::
1019 ..............AlaAspValPheAsnSerGlnThrGlnAsnValGl 1030

3261 GATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAAC 3310
      ::::::          ::::::|||:::   :::|||      :::
1030 yPheSerAspValThrProValIleGluGlnLysAspSerGlyGluLysT 1047

3311 CGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAA 3360
       :::            |||   |||
1047 hrThrTrpThrLeuLysGlyPheAsn..................... 1055

3361 GAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCG 3410
                                    |||:::|||:::|||::
1056 ..................:.............AlaValAlaAsnGlnGl 1061

3411 CGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCG 3460
     ::::||||||
1061 nSerAlaGluLys................................... 1065

3461 CCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAG 3510

1065 ..................................................... 1065

3511 CGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGC 3560
                  :::::::::   ::::::|||   :::::::|||   ||
1066 .........AlaGluAsnPheMetSerAlaGlyTyrLysAsnPheLeuAl 1079

3561 CACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTG 3610
     |   :::|||::::::          :::   :::|||      :::   :
1079 aGluValAsnAsnLeuAsnLysArgMetGlyAspLeuArgAspIleAsnG 1096

3611 CCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAA 3660
     ::|||           |||         ||||||      :::::::::
1096 lyGluAlaGlyAlaTrpAlaArgIleMetSerGlyThrGlySerAlaSer 1112

3661 CACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCG 3710
            |||:::::::::
1113 GlyGlyPheSerAspAsnTyrThrHisVal..................... 1122

3711 CCAAATCGGTATGCAGAAA.....................AACCTCGGCA 3739
      |||:::|||::::::|||                    :::|||   :
1123 .GlnValGlyValAspLysLysHisGluLeuAspGlyLeuAspLeuPheT 1139

3740 GCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAAAACACCTTC 3789
     ::|||               |||:::   :::::::|||:::::::::
1139 hrGly..............PheThrValThrHisThrAspSerSerAla 1150
```

```
3790 GACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGCCGTTTTCGG 3839
1151 SerAlaAspValPheSerGlyLysThrLysSerValGlyAla..... 1164

3840 GCAATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGGGCGCGGGTT 3889
1165 .........................GlyLeuTyrAlaSerAlaMetP 1172

3890 TTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGC 3939
1172 heAspSerGlyAlaTyrIleAspLeuIleGly.............Lys 1183

3940 GTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATT 3989
1184 TyrValHisHis......AspAsnGluTyrThrAlaThrPheAlaGlyLe 1198

3990 CGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGG 4039
1198 uGly.....................ThrArg...............A 1202

4040 ATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTTGCATTCAAC 4089
1202 spTyr.................................SerThrHis 1206

4090 CGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACA 4139
1207 SerTrpTyrAlaGlyAlaGluValGlyTyrArgTyrHisValThrGluAs 1223

4140 CATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGG 4189
1223 pAlaTrpIleGluProGlnAlaGluLeuValTyr...GlySerValSerG 1239

4190 GCAAAGTC..................................... 4197
1239 lyLysGlnPheAlaTrpLysAspGlnGlyMetHisLeuSerMetLysAsp 1255

4198 .....................CGAACACGCGTCAATACCGCCGTATT 4223
1256 LysAspTyrAsnProLeuIleGlyArgThrGlyVal.......... 1267

4224 GGCTCAGGATTTCGGCAAAACCCGCAGTGCG...GAATGGGGCGTAAACG 4270
1268 .......AspValGlyLysSerPheSerGlyLysAspTrpLysValThrA 1282

4271 CC 4272
1282 la 1282
```

seq_name: sp_bacteria:Q93SE3

seq_documentation_block:
ID Q93SE3 PRELIMINARY; PRT; 1363 AA.
AC Q93SE3;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 146.1 KDA PROTEIN.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=4797/97;
RA Zhang W.L., Bielaszewska M., Schmidt H., Karch H.;
RT "Characterisation of a novel pathogenicity island present in LEE-
RT negative Shiga toxin-producing Escherichia coli.";
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AJ278144; CAC39286.1; -.
KW Hypothetical protein.
SQ SEQUENCE 1363 AA; 146130 MW; 2868D6AC45CC9403 CRC64;

alignment_scores:
Quality: 941.00 Length: 1563
Ratio: 1.228 Gaps: 64
Percent Similarity: 49.008 Percent Identity: 25.208

alignment_block:
US-09-303-518D-649 x Q93SE3 ..

Align seg 1/1 to: Q93SE3 from: 1 to: 1363

```
49 AAAACCGGCCGCATCCGCTTCTCGCCTGCTTACTTAGCCATATGCCTGTC 98
29 LysThrGlyArg...ArgLeuMetThrValSerLeuValLeuSerValTh 44

99 GTTCGGCATTCTTCCCCAAGCCTGGGCGGGACACACTTATTTCGGCATCA 148
44 rLeuSerAlaLeuPro...GlyLysAlaSerThrValSerAlaGluIleP 60

149 ACTACCAATACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTC 198
60 roTyrGlnThrPheArgAspPheAlaGluAsnLysGlyValPheThrPro 76

199 GGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAA 248
77 GlyValThrGlyIleGluIleAsnAspAsnAsnGlyAsnLysValGly.. 92

249 ATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACG 298
93 ....ValLeuAspValProMetLeuAspPheSerSerLeuSerArgAspG 108

299 GCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAAC 348
108 lyHisThrThrLeuIleHisProGlyTyrValValSerAlaLysHis... 123

349 GGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCA 398
124 GlyGlyLeuGlnSerValSerSerAlaThrPheGlyTyr......AspGl 138

399 ACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCAGGGA 448
138 n.........IleTyrLysIleValAspAsnAsn.............. 146

449 CTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAA 498
147 ..........LeuAlaGlyIleAspPheSerAlaProArgLeuAsnLys 159

499 TTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGGCG 548
160 LeuValThrGluValIleProAlaAsp.....IleGlnGlyLy 172

549 GAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAG 598
172 sAspLysPheAsnAsnAsnArgTyrThrAlaPheTyrArgAlaGlyValG 189

599 GCAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGT 648
189 lySerGlnTyrIleArg.....................TyrAlaAsnGly 198

649 TCATATCATATTGCAAGTGCGTAT.....TCTTGGCTCGTTGG 686
199 ThrAspLysLeuLeuGlnAlaTyrThrProGluLysAlaTyrLeuThrGl 215

687 TGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAG 736
215 yGlyThrVal.....GlyLysProTyrTyrThrHisTyrAsnG 228

737 GTAGTGAAAAAATTAAACATAGCCCA.....TAT 765
228 lyMetLysMetIleSerAlaAsnProGlyAsnThrPheAspLysAsnGln 244

766 GGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGTT 815
```

```
 245 GlyProLeuAlaSerTyrGlyGlnSerGlyAspSerGlySerProLeuTy 261

 816 TATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAAA 865
     :   :::|||         :::||||||::::::   ||||||
 261 rAlaTrpAspAsnIleAspLysLysTrpValLeuAlaGlyValThrLeuH 278

 866 CGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAA 915
       :::          |||      |||   :::   |||:::
 278 isAsnTyrGlyValLysGlyAlaArgAsnAspTrpLeuLeuIleProHis 294

 916 GATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTA 965
     |||::::::   :::::::::                    :::
 295 AspPheIleSerGlnLysLeuGlnAspAspLeuLysProIleIleValAl 311

 966 CGAACCACGTCAAAAT...GGGAAATACTCTTTTAACGACGATAATAATG 1012
       :::|||:::::|||      ::::::::|||   |||   :::::|
 311 aSerProGluGluAsnIleLeuArgTrpGluPhe...AspArgSerArgG 327

1013 GCACAGGAAAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGA 1062
     ||||||||   :::
 327 lyThrGlyThrLeu.................................... 331

1063 TTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGC 1112
                                              :::
 332 ..............................................SerGl 333

1113 AAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCA 1162
        |||   ::::::      :::|||:::|||||:::          :
 333 nGlyGluLysIlePheSerMetThrGlySerValAsnGly.........A 347

1163 GACTGAATAATGGAGAAAATATTTCCTTTATT...GACGAAGGAAAAGGC 1209
     ::   |||:::|||:::|||:::   |||      :::|||||||||
 347 snAlaAsnThrGlyAsnAsnLeuValPheSerGlyAsnGluGlyLysIle 363

1210 GAATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTT 1259
     ||||||:::      |||:::::::::|||||||||||   |||   ||
 364 GluLeuVal......SerServalGluGlnGlyAlaGlyTyrLeuGlnPh 378

1260 CCAAGGAGATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGCG 1309
     |:::   |||:::||||||      :::|||||:::||||||   ||||
 378 eAspLysAspTyrThrValLeuThrAsnAsnAsnSerThrTrpThrGlyA 395

1310 CGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGC 1359
     |||||:::   ::::::::::::::::::|||   |||   |||||||||
 395 laGlyIleIleValGlyAspGluAlaAsnValLysTrpGlyValAsnGly 411

1360 GTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCA 1409
     :::|||:::|||:::|||   |||:::|||:::|||||||||   ||:::
 412 IleAlaGlyAspAsnLeuHisLysValGlySerGlyThrLeuThrValAs 428

1410 AGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCA 1459
     ::::   |||||||||:::|||:::::::::|||||||||||:::|||:
 428 nGlyHisGlyGluAsnLysGlyGlyLeuLysValGlyAspGlyValValV 445

1460 TTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAA 1509
     ::|||:::||||||   |||   :::   |||:::|||||||||||:::
 445 alLeuGluGlnGlnProAspAlaAsnGlnLysGlnGlnAlaPheSerHis 461

1510 ATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCA 1559
     |||:::::::::|||||||||:::||||||:::||||||:::   |||||
 462 IleAsnIleAlaSerGlyArgAlaThrValLysLeuAsnGlyAlaAsnGl 478

1560 GTTCAACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATT 1609
     |   :::   |||::::::   :::|||:::|||||||||:::|||||||
 478 nValAspAlaAspAsnIleSerTrpGlyTyrArgGlyGlyLysLeuAspL 495

1610 TAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGG 1659
     ||||||||:::::::::::::|||   |||:::|||   :::|||   |||
 495 euAsnGlyTyrAspPheThrPheSerArgLeuGlnAlaAlaAspTyrGly 511

1660 GCGATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTAC 1709
     |||   |||   |||   ||||||      :::|||   ||||||:::::
 512 AlaGluIleSerAsnAspAsnGlnThrAspLysSerIleValThrLeuSe 528
```

```
1710 A..................................................... 1710
     :
 528 rLeuSerProLeuLysAlaGluGluIleAsnValValValAsnAsnIleA 545

1710 ..................................................... 1710

 545 snIleMetGlyGlyThrGlyLysProGlyAspLeuTyrTyrThrThrPhe 561

1711 ...GGCAATAAAGATATTGCTACAACCGGC..................... 1737
       ||||||      :::      ::::::
 562 AspGlyAsnTyrTyrLeuLeuLysSerAsnArgTyrGlySerAlaLeuPh 578

1738 ..........AATAACAACAGC.................TTGGATAGCA 1759
                ||||||:::|||                  |||::::
 578 eGlyAlaLeuAsnAsnGlnSerGluTrpGlnArgLeuGlyLysAspLysG 595

1760 AAAAAGAAATTGCC..................................... 1773
     ::|||   |||:::
 595 luLysAlaIleGlyLeuTyrThrGlnMetLysMetGlnGluSerAlaPro 611

1774 ............TACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAAC 1811
                 |||:::|||                       |||
 612 LeuSerTyrIleTyrHisGly.......................LysIl 620

1812 GAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCC 1861
     :::|||:::   ::::::                |||   :::      |
 620 eThrGlyAsnThrSerValGluIleProLysLeuAlaGlyAsnAspIleL 637

1862 TGCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAAC 1911
     ||   |||:::|||:::::::::::::::::|||:::::::::::   :::
 637 euThrLeuAspGlySerValSerIleSerGlyAspMetSerLysGlnAsp 653

1912 GGCAAACTGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTT 1961
     |||   |||:::|||:::|||:::|||:::   ||||||
 654 GlyAlaLeuIlePheGlnGlyHisProValIleHisAla........... 666

1962 AAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTG. 2010
                                          |||:::   |||
 667 ..................................GlyGlnThrValS 671

2011 .....TGGGACAACGACTGGATCAACCGCACATTTAAAGCGGAAAACTTC 2055
           ::::::||||||   ||||||   |||:::   :::|||:::
 671 erAlaSerGlnSerAspTrpGluAsnArgGluPheSerLeuAsnAsnLeu 687

2056 CAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAAGTGAA 2105
     :::::::::::::::::      :::|||||||||   |||   :::::
 688 AsnLeuAsnAsnAlaAspPheSerLeuSerArgAsn...AlaPheMetAs 703

2106 AGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCAC 2155
     :|||:::   :::      |||::::::   :::::::::|||
 703 nGlyAsnIleArgAlaValAsnGlnSerThrValIleIleGly....... 717

2156 CGCATCAAAGCCACACAATCTGTACA...CGTTCGGACTGGACGGGTCTG 2202
           :::   |||:::   |||   ::::::|||   ||||||
 718 ........GlyAspThrValPheThrAspLysAsnAspGlyThrGly... 730

2203 ACAAATTGTGTCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATT 2252
                  :::   :::|||:::   :::          |||:::
 731 ............AsnAspValIleSerValGluGlyLysSerAlaAlaAl 743

2253 GACTAAGACCGACATCAGCGGCAATGTCGATCTT................. 2286
           ::::::   :::|||::::::   |||
 743 aGlyThrSerSerTyrThrGlyHisIleThrLeuGluGlnLysSerAlaL 760

2287 ................................GCCGATCACGCTCAT 2301
                                     ::::::   :::|||
 760 euAspIleArgAspAsnPheArgGlyGlyValThrSerGluAspSerHis 776

2302 TTAAATCTCACAGGGCTTGCCACACTCAACGGCAATCTTAGTGCA..... 2346
     :::|||:::::::::   ::::::|||   ::::::   |||:::
 777 IleAsnValSerSerSerSerValLeuPheSerAspAlaSerSerPheIl 793
```

```
2347 .AATGGCGATACACGTTATACAGTCAGCCACAACGCCACCCAAAACGGCA 2395
 793 eAsnSerSerLeuAsnIleHisLysGlyGlyAlaLeuThrAlaGlnGlyG 810

2396 ACCTTAGCCTCGTGGGCAATGCCCAAGCAACATTTAATCAAGCCACATTA 2445
 810 lyLeuPheThrSerGlySerIleAspIleGlyAspAlaSerLeuLeuLeu 826

2446 AACGGCAACACATCGGCTTCGGGCAATGCTTCATTT.....AA 2483
 827 ThrGlyThrProValAsnSerAspAspAlaAlaPheLeuProThrIleAs 843

2484 TCTAAGCGAC......CACGCCGTACAAAACGGCAGTCTGACGCTTTCCG 2527
 843 nMetAlaAspGlyGlyPheLysLeuMetSerAspSerSerValLeuLysA 860

2528 GCAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCC 2577
 860 laArgAspGlnAlaSerValVal.............GlyAspIleIle 871

2578 CTAGCCGATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACA 2627
 872 SerAspLysGlnAlaThrIleSerPheGluThr..... 882

2628 AATCAGCGGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGA 2677
 883 ....GluSerGlyLysGluGlyMetLeuSerGluLysAlaSerArg.... 896

2678 CGCTGCCGTCAGGCACGGAATTAGGCAATTTAAACCTTGACAACGCCACC 2727
 897 ..........GlyLeuAlaValGlyLeuLeuSer..... 904

2728 ATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAAACCGG 2777
 905 ...GlyPheAsnThrAlaTyrArg...GlyAlaIleHisAlaProSerAl 919

2778 CAGTGCGACA.....GATGCGCCGCGCCGCCGTTCGC 2809
 919 aSerAlaThrMetAsnAsnThrTrpTrpGlnLeuThrGlyAspSerSerL 936

2810 GCCGTTCGCGCCGTTCCCTATTATCCGTTACA......CCGCCAACTTCG 2853
 936 euArgSerLeuLysAsnThrGlySerMetThrTyrPheThrGlySerAla 952

2854 GTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAAATTGAACGGTCA 2903
 953 AlaAsnLysAlaPheHisThrLeuThrValAsp...GluLeuThrThrAs 968

2904 GGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGACAAAT 2953
 968 nGlyThr.....AlaTyrAlaMetArgThrAsp...L 978

2954 TGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCAACAAT 3003
 978 euLysAsnAlaAspLys.....LeuValValAsnGln 988

3004 ACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAA 3053
 989 Lys.....LeuSerGlyLy 993

3054 AGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAAC 3103
 993 sAspAsnIleLeuLeuValAspPheLeuAsn..... 1003

3104 ACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAG 3153

1003 ..... 1003

3154 TTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGG 3203

1003 ..... 1003

3204 CAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTG 3253

1004 .....LysProThrGlyGluLysLeuAspIleGluLeuValS 1016

3254 ACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTT 3303
1016 erAlaProGlyAsnSerSerLysAspValPheLysGlySerGlu..... 1030

3304 GCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTC...GGCATTATGCA 3350
1031 .....GlnAlaIleGlyPheSerAsnValThrProValIleTh 1043

3351 GGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACC...GCCT 3397
1043 rAlaLysAsp.....AlaGlyAspLysThrThrTrpAsnL 1055

3398 TGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCC 3447
1055 euThrGlyTyrArgMetAlaGluAsnProAlaAlaThrGlnSer..... 1069

3448 CGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCA 3497

1069 ..... 1069

3498 GCCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGA 3547
1070 .....AlaSerGlyLeuAlaSerValGlyTyrL 1079

3548 GTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTA 3597
1079 ysSerPheLeuSerGluValAsnAsnLeu..... 1088

3598 GACCGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCAT 3647

1088 ..... 1088

3648 CCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAAC 3697
1089 .....AsnLysArgM 1092

3698 AAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTC..... 3735
1092 etGlyAspLeuArgAspIleAsnGlyGluAlaGlyAlaTrpAlaArgIle 1108

3736 ......GGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGA 3779
1109 MetSerGlyThrGlySerAlaGlyGlyGlyPheSerAspAsnHisThrHi 1125

3780 AAACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCG 3829
1125 s......ValGlnValGlyValAspLysLysHisGluLeuAspGlyLeuA 1140

3830 CCGTTTTCGGGCAATACGGCATC..... 3852
1140 spLeuPheThrGlyPheThrValThrHisThrAspSerSerAlaSerAla 1156

3853 GACAGGTTCTAC.....ATCGGCATCAGCGCGGG 3881
1157 AspAlaPheLysGlyLysThrLysSerValGlyAlaGlyLeuTyrAlaSe 1173

3882 CGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGGCAAAATCC 3931
1173 rAlaMetPheAspSerGlyAlaTyrIleAspLeuIleGly..... 1186

3932 GCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTC 3981
1187 .....LysTyrValHisHis.....AspAsnGluTyrThrAlaThrPhe 1199

3982 GGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCA 4031
1200 AlaGlyLeuGlyIleArg..... 1205

4032 AAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTTG 4081
```

```
 1206 .......AspTyr.................................S 1208

 4082 CATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCG 4131
      ::   :::   :::   ||||||   :::|||   |||
 1208 erThrHisSerTrpTyrAlaGlyAlaGluAlaGlyTyrArgCysHisVal 1224

 4132 GCGCAACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGC 4181
      ::::::         |||   |||      :::|||   |||      ||
 1225 ThrGluAspThrTrpIleGluProGlnAlaGluLeuValTyr...GlyAl 1240

 4182 CGCTTCGGGCAAA.................................G 4195
      |:::||||||||||                                 :
 1240 aValSerGlyLysGlnPheAlaTrpLysAspGlnGlyMetHisLeuSerM 1257

 4196 TCCGAACACGCGTCAATACCGCCGTATTGGCTCAG.........GATTTC 4236
      :::::   |||      :::   ::::::::::::        |||
 1257 etLysAspArgAspTyrAsnProLeuIleGlyArgThrGlyValAspVal 1273

 4237 GGCAAAACCCGCAGTGCG...GAATGGGGCGTAAACGCC 4272
      ||||||:::   |||:::   :::|||   |||:::|||
 1274 GlyLysSerPheSerGlyLysAspTrpLysValThrAla 1286

seq_name: sp_bacteria:Q51171

seq_documentation_block:
ID   Q51171      PRELIMINARY;      PRT;   508 AA.
AC   Q51171;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NG117;
RX   MEDLINE=95302961; PubMed=7783620;
RA   Lomholt H., Poulsen K., Mogens K.;
RT   "Comparative characterization of the iga gene encoding IgA1 protease
RT   in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
RT   influenzae.";
RL   Mol. Microbiol. 15:495-506(1995).
DR   EMBL; X82483; CAA57866.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     508    508
SQ   SEQUENCE   508 AA;  56035 MW;  792853E8187E8D06 CRC64;


alignment_scores:
            Quality:  890.00              Length:    563
              Ratio:  2.500                Gaps:     21
 Percent Similarity:  63.233   Percent Identity: 38.721

alignment_block:
 US-09-303-518D-649 x Q51171   ..

 Align seg 1/1  to: Q51171  from: 1  to: 508

     202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
         |||   |||::::::|||   ||||||:::|||:::   :::|||:::::
       1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

     252 AATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAAC...G 298
         :::::::::::::|||||||||||||||||||||||||::::::   |||
      17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnLysA 34

     299 GCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT... 345
```

```
         :::|||:::::::|||      ||||||   |||||||||   |||
  34 rgIleAlaThrValValAspProGlnTyrAlaValSerValLysHisAla 50

 346 ......................AACGGCGGCTATAACAACGTTGATTT 371
                                 |||   ||||||:::   :::
  51 LysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAsnAs 67

 372 TGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATTG 421
       |||:::   :::|||                          |||:::::::|
  67 pValAlaAspLysGluAsn.................GluTyrArgValV 78

 422 TGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGC 471
     ||:::::::||||||||||:::               |||         |||
  78 alGluGlnAsnAsnTyrGluPro............HisLysAlaTrpGly 90

 472 ......................GATTATCATATGCCGCGTTTGCATAA 497
                            ||||||:::|||   |||::::::||
  91 AlaSerAsnLeuGlyArgLeuGluAspTyrAsnMetAlaArgPheAsnLy 107

 498 ATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGG. 546
     ||||||||||::::::   |||:::      |||:::          |||
 107 sPheValThrGluValAlaProIleAlaProThrAspAlaGlyGlyGlyL 124

 547 ..CGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGG 594
           |||   |||:::|||::::::   :::   ||||||:::|||
 124 euAspThrTyrLysAspLysAsnArgPheSerSerPheValArgValGly 140

 595 GCAGGCAGGCAA......TATTGGCGATCTGATGAAGATGAGCCCAATAA 638
     ||||||||||||      |||   ::::::         :::   :::||
 141 AlaGlyArgGlnLeuValTyrGluLysGlyAlaTyrHisGlnGluGlyAs 157

 639 CCGCGAAAGTTCATAT.........CATATTGCAAGTGCGTATTCTTGGC 679
     |   |||:::::::|||           :::::::::||||||   :::
 157 n...GluLysGlyTyrAspLeuArgAspLeuSerGlnAlaTyrArgTyrA 173

 680 TCGTTGGTGGC.................................AAT 693
       :::::::|||                                  |||
 173 laIleAlaGlyThrProTyrLysAspIleAsnIleAspGlnThrMetAsn 189

 694 ACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGA 743
     |||   :::      |||   |||:::   :::::::::   ::::::||
 190 ThrGluGlyLeuIleGlyPheGlyAsnHisAsnThrHisTyrSerAlaGl 206

 744 AAAAATTAAACATAGC...CCATATGGTTTTTTACCAACAGGAGGCTCAT 790
     |:::::::|||:::::                |||   :::   |||   :
 206 uGluLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrGlyValL 223

 791 TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
     ::|||||||||||||||||||:::|||   :::|||   ||||||:::::
 223 euGlyAspSerGlySerProLeuPheAlaPheAspLysGlnLysAsnGln 239

 841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAG 890
     |||::::::   |||:::   :::            |||   ||||||
 240 TrpValPheLeuGlyThrTyrAspTyrTrpAlaGlyTyr...GlyLys.. 254

 891 CAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTG 940
                         |||:::|||         :::|||:::
 255 .....................LysSerTrpGlnGluTrpAsnIleTyr. 263

 941 CTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATAC 990
                                                 |||
 264 ..........................................LysLys 265

 991 TCTTTTAACGAC..............GATAATAATGGCACAGGAAAAAT 1025
     :::|||   |||               ||||||   ||||||   |||
 266 GluPheAlaAspLysIleLysGlnArgAspAsnAlaGlyThrIleLysGl 282

1026 CAATGCCAAACATGAACACAAT.........TCTCTGCCTAATAGATTAA 1066
      |||   :::|||:::   |||        :::      :::::::::
 282 yAsnArgGluHisHisTrpAsnIleThrPheGlyThrAsnSerLysIleG 299

1067 AAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGA 1116
       :::   :::|||:::|||   |||               :::   |||
```

```
 299 lySerThrAlaValArgLeuAlaAsn.............AsnGluArg 310

1117 GAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACT 1166
      :::
 311 Asp.....................................Al 312

1167 GAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGA 1216
      ||||||||||:::|||:::::::|||      |||   :::|||   |||:
 312 aAsnAsnGlyGlnAsnValThrPhe......GluAspAsnGlyThrLeuV 327

1217 TACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGA 1266
      ::|||   :::|||||||||||||||||||||||||||:::|||:::||| 
 327 alLeuAspGlnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGly 343

1267 GATTTTACGGTCTCGCCTGAAAATAACGAA...ACTTGGCAAGGCGCGGG 1313
      |||:::||||||:::      ||||||:::   ||||||   ||||||||
 344 AspTyrThrValLysGlyAlaAsnAsnAspIleThrTrpLeuGlyAlaGl 360

1314 CGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGG 1363
      |:::   :::::::::   :::   |||:::|||:::|||:::::::
 360 yIleAspValAlaAspGlyLysLysValValTrpGlnValLysAsnProA 377

1364 CAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCC 1413
       :::||||||||||:::|||||||||||||||||||||||:::::::::::
 377 snGlyAspArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGly 393

1414 AAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTT 1463
       |||   ||||||||:::::::::|||||||||||||||||||||||
 394 ThrGlyValAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLe 410

1464 GGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCG 1513
      |:::||||||||||||          |||   |||||||||||||:::::|
 410 uAsnGlnGlnAlaAspAlaAspLysLysValGlnAlaPheSerGlnValG 427

1514 GCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTC 1563
      ||:::|||||||||||||||||:::   ||||||:::::|||||:::
 427 lyIleValSerGlyArgGlyThrLeuValLeuAsnSerSerAsnGlnIle 443

1564 AACCCCGACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAA 1613
      |||||||||:::||||||||||||||||||||||||||||||||||   ||
 444 AsnProAspAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAs 460

1614 CGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGA 1663
      ||||:::::::|||:::|||:::::::|||:::|||:::||||||||||||
 460 nGlyAsnAspLeuThrPheGluHisIleArgAsnValAspGluGlyAlaA 477

1664 TGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGC 1713
       |||||||||||||||   |||:::   ||||||:::|||:::||||||
 477 rgIleValAsnHisAsnThrAspArgAlaSerThrIleThrLeuThrGly 493

1714 AATAAAGATATTGCTACAACCGGCAATAACAACAGCTTG 1752
       |||::::::   |||:::   :::|||   |||:::
 494 ...LysSerLeuIleThrAlaProGlnAsnLeuSerVal 505

seq_name: sp_bacteria:Q51165

seq_documentation_block:
ID   Q51165      PRELIMINARY;      PRT;   496 AA.
AC   Q51165;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NG44/76;
RX   MEDLINE=95302961; PubMed=7783620;
RA   Lomholt H., Poulsen K., Mogens K.;
RT   "Comparative characterization of the iga gene encoding IgA1 protease
RT   in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
RT   influenzae.";
RL   Mol. Microbiol. 15:495-506(1995).
DR   EMBL; X82481; CAA57864.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     496    496
SQ   SEQUENCE   496 AA;  54645 MW;  AED2478E59F3CB84 CRC64;


alignment_scores:
            Quality:  883.50              Length:    546
              Ratio:  2.461                Gaps:     18
 Percent Similarity:  65.751   Percent Identity: 38.462

alignment_block:
 US-09-303-518D-649 x Q51165   ..

 Align seq 1/1  to: Q51165  from: 1  to: 496

 202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
      |||   |||::::::|||   ||||||:::|||:::   :::|||:::::
   1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

 252 AATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAAC...G 298
      ::::::::::::|||||||||||||||||||||||:::::   |||
  17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnArgA 34

 299 GCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAAC 348
       :::   :::::::::      ||||||   |||||||||   |||
  34 rgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50

 349 GGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGAT.. 396
       |||   :::         :::|||   ::::::   :::   |||
  51 LysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAspVa 67

 397 ....CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAG 442
          :::   :::        |||:::   |||   :::|||:::|||:::
  67 lSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAspTyrGluP 84

 443 CAGGGACTAAAGGCCATCCTTATGGCGGC.............GATTAT 477
       :::   :::   |||   :::|||:::                ||||||
  84 roAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuGluAspTyr 99

 478 CATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAAT 527
      :::|||   ||||||:::|||||||||||:::::   |||:::
 100 AsnMetAlaArgLeuAsnLysPheValThrGluValAlaProIleAlaPr 116

 528 GACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAATAATTACC 574
      ||||||         |||   :::   |||   |||:::|||:::::::
 116 oThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysAsnArgPheS 133

 575 CTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAA 624
       :::   ||||||:::|||||||||   |||:::
 133 erGluPheValArgValGlyAlaGlyThrGlnPhe.............. 144

 625 GATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTC 674
       |||   |||:::|||   :::      :::::::::   ||||||
 145 ...GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArgAlaTyrAr 160

 675 TTGGCTCGTTGGTGGCAATACCTTT.............GCACAAAATG 709
       :::   ::::::|||:::   :::                ::::::|||
 160 gTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValThrSerAsnL 177

 710 GATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGC 759
       :::      |||   :::::::::|||:::::::::      ||||||
 177 euAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLysHisHisSer 193

 760 CCA.....................TATGGTTTTTTACCAACAGGAGG 785
```

```
     |||                                     :::   |||   :::   ::
 194 ProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuThrAsnTyrAl 210

 786 CTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGC 835
     :   :::||||||||||||||||||:::|||  ||||||  |||::::
 210 aValLeuGlyAspSerGlySerProLeuPheAlaTyrAspLysGlnGluL 227

 836 AAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGA 885
     :::::|||:::::  |||:::    :::             |||
 227 ysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGlyTyr...... 241

 886 AAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG...TTCTATGATGA 932
     ::::::|||::::::|||            :::|||   :::|||   ::
 242 GlnLysAsnSerTrpGln............GluTrpAsnIleTyrLysLy 254

 933 AATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATG 982
     :   ||||||
 254 sGluPheAlaAspLysIleLys............................ 261

 983 GGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAATCAATGCC 1032
                   :::  ||||||  ||||||   |||   |||:::
 262 ..............GlnArgAspAsnAlaGlyThrIleLysGlyAsnGly 273

1033 AAACATGAACACAAT.........TCTCTGCCTAATAGATTAAAAACACG 1073
     :::|||:::   |||          :::       :::::::::   :::
 274 GluHisHisTrpAsnIleThrPheGlyThrAsnSerLysIleGlySerTh 290

1074 AACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTG 1123
     :::|||:::|||
 290 rAlaValArgLeu..................................... 294

1124 TTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAAT 1173
                 ||||||:::  ::::::             ||||||
 295 ...........AlaGlyAsnGluLysAsp............AlaAsnAsn 303

1174 GGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTAC 1223
     |||:::|||::::::|||       |||   :::|||  |||:::|||
 304 GlyGlnAsnValThrPhe......GluAspAsnGlyThrLeuValLeuAs 318

1224 CAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTA 1273
     :::||||||||||||||||||||||||||||:::|||:::||||||:::|
 318 pGlnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrT 335

1274 CGGTCTCGCCTGAAAATAACGAA...ACTTGGCAAGGCGCGGGCGTTCAT 1320
     ||||||:::     |||||||:::  ||||||  |||||||||:::
 335 hrValLysGlyIleAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAsp 351

1321 ATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGA 1370
     :::::::::   :::   |||:::|||:::|||::::::      :::||
 352 ValThrAspGlyLysLysValValTrpGlnValLysAsnProAsnGlyAs 368

1371 CCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGG 1420
     ||||||||:::|||||||||||||||||||||||:::::::::::::   |||
 368 pArgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyV 385

1421 AAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAG 1470
     ||||||||:::::::::||||||||||||||||||||||||:::|||
 385 alAsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGln 401

1471 CAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGT 1520
     |||||||||         |||   |||||||||||||::::::|||:::||
 402 GlnAlaAspAlaAspLysLysValGlnAlaPheSerGlnValGlyIleVa 418

1521 CAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCG 1570
     |||||||||||||||:::   ||||||::::::||||||:::||||||
 418 lSerGlyArgGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProA 435

1571 ACAAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCAT 1620
     ||:::||||||||||||||||||||||||||||||||||   |||||||:::
 435 spAsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsn 451

1621 TCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGT 1670
     :::|||:::|||:::::::|||:::|||:::|||||||||||   |||||
 452 AspLeuThrPheGluHisIleArgAsnValAspGluGlyAlaArgIleVa 468

1671 CAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAG 1720
     |||||||||     :::          ||||||:::|||:::||||||   |||:
 468 lAsnHisAsnThrSerHisAlaSerThrIleThrLeuThrGly...LysS 484

1721 ATATTGCTACAACCGGCAATAACAACAGCTTGGATAGC 1758
     :::::     |||:::   |||:::    |||:::   |||
 484 erLeuIleThrAsnProAsnSerLeuSerValHisSer 496
```

seq_name: sp_bacteria:Q51163

```
seq_documentation_block:
ID   Q51163       PRELIMINARY;      PRT;   496 AA.
AC   Q51163;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=81139;
RX   MEDLINE=95302961; PubMed=7783620;
RA   Lomholt H., Poulsen K., Mogens K.;
RT   "Comparative characterization of the iga gene encoding IgA1 protease
RT   in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
RT   influenzae.";
RL   Mol. Microbiol. 15:495-506(1995).
DR   EMBL; X82477; CAA57860.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     496    496
SQ   SEQUENCE   496 AA;  54766 MW;  BF4C7A7012FF4D99 CRC64;

alignment_scores:
            Quality:  873.00              Length:     534
              Ratio:  2.452                 Gaps:      20
 Percent Similarity:  66.667    Percent Identity:  39.513

alignment_block:
 US-09-303-518D-649 x Q51163  ..

 Align seg 1/1  to: Q51163  from: 1  to: 496

 202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
     |||   |||::::::|||   ||||||:::|||:::   :::|||:::::
   1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

 252 AATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAAC...G 298
     ::::::::::::||||||||||||||||||||||||::::::   |||
  17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnArgA 34

 299 GCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAAC 348
     :::   :::::::::        ||||||   |||||||||   |||
  34 rgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50

 349 GGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGAT.. 396
     |||   :::           :::|||   ::::::   :::   |||
  51 LysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAspVa 67

 397 ....CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAG 442
       :::   :::        |||:::   |||   :::|||:::|||:::
  67 lSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAspTyrGluP 84
```

```
443 CAGGGACTAAAGGCCATCCTTATGGCGGC.............GATTAT 477
     :::   :::   |||   :::|||:::               ||||||
 84 roAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuGluAspTyr 99

478 CATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAAT 527
    :::|||   ||||||:::|||||||||||||::::::   |||:::
100 AsnMetAlaArgLeuAsnLysPheValThrGluValAlaProIleAlaPr 116

528 GACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAATAATTACC 574
     ||||||          |||   :::   |||   |||:::|||::::::
116 oThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysAsnArgPheS 133

575 CTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAA 624
      :::   ||||||:::|||||||||   |||:::
133 erGluPheValArgValGlyAlaGlyThrGlnPhe.............. 144

625 GATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTC 674
       |||   |||:::|||   :::      :::::::::   ||||||
145 ...GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArgAlaTyrAr 160

675 TTGGCTCGTTGGTGGCAATACCTTT.............GCACAAAATG 709
     :::   ::::::|||:::   :::                :::::|||
160 gTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValThrSerAsnL 177

710 GATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGC 759
      :::      |||   ::::::::::|||:::::::::      ||||||
177 euAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLysHisHisSer 193

760 CCA.....................TATGGTTTTTTACCAACAGGAGG 785
    |||                          :::   |||   :::   ::
194 ProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuThrAsnTyrAl 210

786 CTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGC 835
    :   :::||||||||||||||||||:::|||   ||||||   |||::::
210 aValLeuGlyAspSerGlySerProLeuPheAlaTyrAspLysGlnGluL 227

836 AAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGA 885
    :::::|||::::::   |||:::   :::            |||
227 ysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGlyTyr..... 241

886 AAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG...TTCTATGATGA 932
    :::::::|||:::::::|||          :::|||   :::|||   .::
242 GlnLysAsnSerTrpGln............GluTrpAsnIleTyrLysLy 254

933 AATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATG 982
    :   ||||||                        |||   :::|||:::
254 sGluPheAlaAsp.....................GluIleLysGlnArg. 263

983 GGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAATCAATGCC 1032
                      ||||||   ||||||   |||   |||:::
264 ...................AspAsnAlaGlyThrIleLysGlyAsnGly 273

1033 AAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCA 1082
     :::|||:::   ::::::                |||   |||
274 GluHisHisTrpLysThr.................ThrGlyThr..... 282

1083 ATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATG 1132
           |||      :::::::::|||                    |
283 .......AsnSerHisIleGlySerThr....................A 290

1133 CTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAAT 1182
     ||:::      :::|||:::   |||         |||||||||:::|||
290 laValArgLeuAlaAsnAsnGluArgAsp...AlaAsnAsnGlyGlnAsn 305

1183 ATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAACAT 1232
     ::::::|||      |||:::::::|||   |||:::|||   :::||||||
306 ValThrPhe......GluAsnAsnGlyThrLeuValLeuAspGlnAsnIl 320

1233 CAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCTCGC 1282
     ||||||||||||||||||||||:::|||:::||||||:::||||||:::
320 eAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrThrValLysG 337

1283 CTGAAAATAACGAA...ACTTGGCAAGGCGCGGGCGTTCATATCAGTGAA 1329
```

```
          ||||||:::   ||||||   |||||||||:::   :::::::::
337 lyIleAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspValAlaAsp 353

1330 GACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTC 1379
       :::   |||:::|||:::|||::::::      :::||||||||::
354 GlyLysLysValValTrpGlnValLysAsnProAsnGlyAspArgLeuAl 370

1380 CAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAG 1429
     :||||||||||||||||||||:::::::::::   |||   |||||||
370 aLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyValAsnGlnG 387

1430 GCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGAC 1479
     ||:::::::::|||||||||||||||||||||||:::|||||||||||
387 lyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnGlnAlaAsp 403

1480 GATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAG 1529
               |||   ||||||||||||:::::::|||:::||||||||||
404 AlaAspLysLysValGlnAlaPheSerGlnValGlyIleValSerGlyAr 420

1530 GGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCT 1579
     ||||||||:::  ||||||:::::|||||:::||||||||||:::||||
420 gGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeuT 437

1580 ATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCG 1629
     |||||||||||||||||||||||||||||   ||||||:::::|||:::
437 yrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThr 453

1630 TTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAA 1679
     |||::::::|||:::|||:::||||||||||   ||||||||||||||
454 PheGluHisIleArgAsnValAspGluGlyAlaArgIleValAsnHisAs 470

1680 TCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTA 1729
     |            ||||||:::|||:::||||||   |||::::::   |
470 nThrGlyHisAlaSerThrIleThrLeuThrGly...LysSerLeuIleT 486

1730 CA 1731
     ||
486 hr 486
```

seq_name: sp_bacteria:Q57386

seq_documentation_block:

```
ID   Q57386      PRELIMINARY;      PRT;   496 AA.
AC   Q57386;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SM1027;
RX   MEDLINE=95302961; PubMed=7783620;
RA   Lomholt H., Poulsen K., Mogens K.;
RT   "Comparative characterization of the iga gene encoding IgA1 protease
RT   in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilus
RT   influenzae.";
RL   Mol. Microbiol. 15:495-506(1995).
DR   EMBL; X82472; CAA57855.1; -.
DR   EMBL; X82470; CAA57853.1; -.
DR   EMBL; X82471; CAA57854.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     496    496
SQ   SEQUENCE   496 AA;  54634 MW;  4A4F91C0C44AF866 CRC64;
```

```
alignment_scores:
            Quality:  872.00              Length:    541
              Ratio:  2.422                 Gaps:     19
 Percent Similarity:  66.543   Percent Identity: 38.817

alignment_block:
 US-09-303-518D-649 x Q57386   ..

 Align seg 1/1  to: Q57386  from: 1  to: 496

     202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
         |||   |||::::::|||   |||||||:::|||:::   :::|||:::::
       1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

     252 AATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAAC...G 298
         :::::::::::::|||||||||||||||||||||::::::   |||
      17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnArgA 34

     299 GCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAAC 348
         :::   :::::::::      ||||||   |||||||||   |||
      34 rgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50

     349 GGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGAT.. 396
         |||   :::         :::|||   ::::::   :::   |||
      51 LysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAspVa 67

     397 ....CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAG 442
             :::   :::      |||:::   |||   :::|||:::|||:::
      67 lSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAspTyrGluP 84

     443 CAGGGACTAAAGGCCATCCTTATGGCGGC.............GATTAT 477
         :::   :::   |||   :::|||:::                |||||
      84 roAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuGluAspTyr 99

     478 CATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAAT 527
         :::|||   ||||||:::|||||||||||||:::::   |||:::
     100 AsnMetAlaArgLeuAsnLysPheValThrGluValAlaProIleAlaPr 116

     528 GACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAATAATTACC 574
         ||||||          |||   :::   |||   |||:::|||:::::
     116 oThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysAsnArgPheS 133

     575 CTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAA 624
         :::   ||||||:::|||||||||   |||:::
     133 erGluPheValArgValGlyAlaGlyThrGlnPhe..............  144

     625 GATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTC 674
            |||   |||:::|||   :::      :::::::::   ||||||
     145 ...GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArgAlaTyrAr 160

     675 TTGGCTCGTTGGTGGCAATACCTTT..............GCACAAAATG 709
         :::   ::::::|||:::   :::                :::::::|||
     160 gTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValThrSerAsnL 177

     710 GATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGC 759
           :::      |||   :::::::::|||:::::::::      ||||||
     177 euAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLysHisHisSer 193

     760 CCA.......................TATGGTTTTTTACCAACAGGAGG 785
         |||                       :::   |||   :::   ::
     194 ProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuThrAsnTyrAl 210

     786 CTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGC 835
         :   :::||||||||||||||||||:::|||   ||||||   |||::::
     210 aValLeuGlyAspSerGlySerProLeuPheAlaTyrAspLysGlnGluL 227

     836 AAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGA 885
         :::::|||::::::   |||:::   :::          |||
     227 ysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGlyTyr...... 241

     886 AAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG...TTCTATGATGA 932
         ::::::|||::::::|||            :::|||   :::|||   ::
     242 GlnLysAsnSerTrpGln............GluTrpAsnIleTyrLysLy 254

     933 AATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATG 982
         :   ||||||                          |||   :::|||:::
     254 sGluPheAlaAsp.....................GluIleLysGlnArg. 263

     983 GGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAATCAATGCC 1032
                         ||||||   ||||||   |||      :::
     264 ...................AspAsnAlaGlyThrIleLysGlyTyrGly 273

    1033 AAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCA 1082
         :::|||:::   ::::::                  |||   |||
     274 GluHisHisTrpLysThr.................ThrGlyThr..... 282

    1083 ATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATG 1132
                |||       :::::::::||||||   :::   :::   :::
     283 .......AsnSerHisIleGlySerThrAlaValArgLeuAlaGlyAsnG 297

    1133 CTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAAT 1182
             |||:::                          |||||||||:::|||
     297 luArgGlyAla.......................AsnAsnGlyGlnAsn 305

    1183 ATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAACAT 1232
         ::::::|||      |||:::::|||   |||:::|||   :::|||||
     306 ValThrPhe......GluAsnAsnGlyThrLeuValLeuAspGlnAsnIl 320

    1233 CAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCTCGC 1282
         ||||||||||||||||||||||:::|||:::||||||:::|||||||:::
     320 eAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrThrValLysG 337

    1283 CTGAAAATAACGAA...ACTTGGCAAGGCGCGGGCGTTCATATCAGTGAA 1329
             ||||||      ||||||   |||||||||:::   :::::::::
     337 lyAlaAsnAsnGlyIleThrTrpLeuGlyAlaGlyIleAspValAlaAsp 353

    1330 GACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTC 1379
            :::   |||:::|||:::|||::::::      :::||||||||::
     354 GlyLysLysValValTrpGlnValLysAsnProAsnGlyAspArgLeuAl 370

    1380 CAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAG 1429
         :|||||||||||||||||||||:::::::::::::   |||   |||||||
     370 aLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyValAsnGlnG 387

    1430 GCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGAC 1479
         ||:::::::::|||||||||||||||||||||:::|||||||||||||
     387 lyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnGlnAlaAsp 403

    1480 GATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAG 1529
                |||   |||||||||||:::::::|||:::|||||||||||
     404 AlaAspLysLysValGlnAlaPheSerGlnValGlyIleValSerGlyAr 420

    1530 GGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCT 1579
         |||||||:::   |||||||::   :::|||:::||||||:::::||||
     420 gGlyThrLeuValLeuAsnSerProAspGlnIleAsnProAsnAsnLeuT 437

    1580 ATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCG 1629
         ||||||||||||||||||||||||||||   ||||||:::::|||:::
     437 yrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThr 453

    1630 TTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAA 1679
         |||:::::::|||:::|||:::||||||||||||   ||||||||||||
     454 PheGluHisIleArgAsnValAspGluGlyAlaArgIleValAsnHisAs 470

    1680 TCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTA 1729
         |   |||:::   ||||||:::|||:::||||||   |||::::::   |
     470 nThrAspArgAlaSerThrIleThrLeuThrGly...LysSerLeuIleT 486

    1730 CAACCGGCAATAACAACAGCTTG 1752
         ||:::   :::|||   |||:::
     486 hrAlaProGlnAsnLeuSerVal 493

seq_name: sp_bacteria:Q51166

seq_documentation_block:
```

```
ID   Q51166      PRELIMINARY;      PRT;   496 AA.
AC   Q51166;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SM1894;
RX   MEDLINE=95302961; PubMed=7783620;
RA   Lomholt H., Poulsen K., Mogens K.;
RT   "Comparative characterization of the iga gene encoding IgA1 protease
RT   in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
RT   influenzae.";
RL   Mol. Microbiol. 15:495-506(1995).
DR   EMBL; X82476; CAA57859.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     496    496
SQ   SEQUENCE   496 AA;  54725 MW;  A1B997BB7F0CEF82 CRC64;


alignment_scores:
            Quality:  870.00              Length:    543
              Ratio:  2.403                Gaps:     19
 Percent Similarity:  66.667   Percent Identity: 38.674

alignment_block:
 US-09-303-518D-649 x Q51166   ..

 Align seg 1/1  to: Q51166  from: 1  to: 496

     202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
         |||   |||:::::::|||   |||||||:::|||:::   :::|||:::::
       1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

     252 AATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAAC...G 298
         :::::::::::::|||||||||||||||||||||||:::::   |||
      17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnArgA 34

     299 GCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAAC 348
           :::   :::::::::       |||||||   |||||||||   |||
      34 rgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50

     349 GGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGAT.. 396
           |||   :::          :::|||   :::::::   :::   |||
      51 LysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAspVa 67

     397 ....CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAG 442
             :::   :::        |||:::   |||   :::|||:::|||:::
      67 lSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAspTyrGluP 84

     443 CAGGGACTAAAGGCCATCCTTATGGCGGC..............GATTAT 477
           :::   :::   |||   :::|||:::                |||||
      84 roAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuGluAspTyr 99

     478 CATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAAT 527
         :::|||   ||||||:::|||||||||||||||:::::   |||:::
     100 AsnMetAlaArgLeuAsnLysPheValThrGluValAlaProIleAlaPr 116

     528 GACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAATAATTACC 574
           ||||||         |||   :::   |||   |||:::|||::::::
     116 oThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysAsnArgPheS 133

     575 CTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAA 624
         :::   ||||||:::|||||||||   |||:::
     133 erGluPheValArgValGlyAlaGlyThrGlnPhe.............. 144

     625 GATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTC 674
           |||   |||:::|||   :::      :::::::::   ||||||
     145 ...GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArgAlaTyrAr 160

     675 TTGGCTCGTTGGTGGCAATACCTTT.............GCACAAAATG 709
         :::   ::::::|||:::   :::               :::::::|||
     160 gTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValThrSerAsnL 177

     710 GATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGC 759
           :::      |||   :::::::::|||:::::::::      ||||||
     177 euAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLysHisHisSer 193

     760 CCA.....................TATGGTTTTTTACCAACAGGAGG 785
         |||                     :::   |||  :::   ::
     194 ProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuThrAsnTyrAl 210

     786 CTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGC 835
         :   :::|||||||||||||||||:::|||   ||||||   |||::::
     210 aValLeuGlyAspSerGlySerProLeuPheAlaTyrAspLysGlnGluL 227

     836 AAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGA 885
         :::::|||:::::   |||:::   :::          |||
     227 ysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGlyTyr..... 241

     886 AAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG...TTCTATGATGA 932
         ::::::|||:::::::|||          :::|||   :::|||   ::
     242 GlnLysAsnSerTrpGln.............GluTrpAsnIleTyrLysLy 254

     933 AATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATG 982
         :   ||||||                    |||   :::|||:::
     254 sGluPheAlaAsp.....................GluIleLysGlnArg. 263

     983 GGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAATCAATGCC 1032
                                 ||||||   ||||||   |||      :::
     264 ...................AspAsnAlaGlyThrIleLysGlyTyrGly 273

    1033 AAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCA 1082
         :::|||:::   ::::::                    |||   |||
     274 GluHisHisTrpLysThr.................ThrGlyThr..... 282

    1083 ATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATG 1132
              |||      :::::::::||||||   :::   :::   :::
     283 .......AsnSerHisIleGlySerThrAlaValArgLeuAlaGlyAsnG 297

    1133 CTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAAT 1182
               |||:::                          ||||||||||:::|||
     297 luArgGlyAla.....................AsnAsnGlyGlnAsn 305

    1183 ATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAACAT 1232
         ::::::|||      |||:::::::|||   |||:::|||   :::||||||
     306 ValThrPhe......GluAsnAsnGlyThrLeuValLeuAspGlnAsnIl 320

    1233 CAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCTCGC 1282
         ||||||||||||||||||||||:::|||:::||||||:::||||||:::
     320 eAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrThrValLysG 337

    1283 CTGAAAATAACGAA...ACTTGGCAAGGCGCGGGCGTTCATATCAGTGAA 1329
               ||||||:::   ||||||   |||||||||:::   :::::::::
     337 lyIleAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspValAlaAsp 353

    1330 GACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTC 1379
           :::   |||:::|||:::|||::::::      :::||||||||||::
     354 GlyLysLysValValTrpGlnValLysAsnProAsnGlyAspArgLeuAl 370

    1380 CAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAG 1429
         :||||||||||||||||||||:::::::::::   |||   |||||||
     370 aLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyValAsnGlnG 387

    1430 GCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGAC 1479
         ||:::::::::|||||||||||||||||||||||||:::|||::||||||
```

```
 387 lyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnLysAlaAsp 403

1480 GATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAG 1529
     ::::::   |||   ||||||||||||:::::|||:::|||||||||||
 404 SerAsnGlnLysValGlnAlaPheSerGlnValGlyIleValSerGlyAr 420

1530 GGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCT 1579
     |   |||:::   ||||||:::::::|||||||:::|||||||||:::||||
 420 gArgThrLeuValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeuT 437

1580 ATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCG 1629
     |||||||||||||||||||||||||||||   ||||||:::::::|||:::
 437 yrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThr 453

1630 TTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAA 1679
     |||::::::|||:::|||:::||||||||||||   ||||||||||||||
 454 PheGluHisIleArgAsnValAspGluGlyAlaArgIleValAsnHisAs 470

1680 TCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTA 1729
     |            ||||||:::|||:::||||||   |||::::::   |
 470 nThrGlyHisThrSerThrIleThrLeuThrGly...LysSerLeuIleT 486

1730 CAACCGGCAATAACAACAGCTTGGATAGC 1758
     ||:::   |||:::   |||:::   |||
 486 hrAsnProAsnSerLeuSerValHisSer 495
```

```
seq_name: sp_bacteria:Q51167

seq_documentation_block:
ID   Q51167      PRELIMINARY;      PRT;   508 AA.
AC   Q51167;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NGC16;
RX   MEDLINE=95302961; PubMed=7783620;
RA   Lomholt H., Poulsen K., Mogens K.;
RT   "Comparative characterization of the iga gene encoding IgA1 protease
RT   in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
RT   influenzae.";
RL   Mol. Microbiol. 15:495-506(1995).
DR   EMBL; X82485; CAA57868.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     508    508
SQ   SEQUENCE   508 AA;  55879 MW;  9A88D76477BA25C0 CRC64;


alignment_scores:
            Quality:  870.00              Length:    553
              Ratio:  2.465                Gaps:     22
 Percent Similarity:  63.834   Percent Identity: 39.060

alignment_block:
 US-09-303-518D-649 x Q51167   ..

 Align seg 1/1  to: Q51167  from: 1  to: 508

  202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
      |||   |||::::::|||   ||||||:::|||:::   :::|||:::::
    1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

  252 AATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAAC...G 298
```

```
      :::::::::::::|||||||||||||||||||||:::::   |||
   17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnLysA 34

  299 GCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT... 345
        :::|||:::::::|||      ||||||   |||||||||   |||
   34 rgIleAlaThrValValAspProGlnTyrAlaValSerValLysHisAla 50

  346 ........................AACGGCGGCTATAACAACGTTGATTT 371
                                  |||   ||||||:::   :::
   51 LysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAsnAs 67

  372 TGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATTG 421
        |||:::   :::|||                        |||:::::::|
   67 pValAlaAspLysGluAsn.................GluTyrArgValV 78

  422 TGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGC 471
      ||:::::::|||||||||:::                |||         |||
   78 alGluGlnAsnAsnTyrGluPro............HisLysAlaTrpGly 90

  472 ........................GATTATCATATGCCGCGTTTGCATAA 497
                                  |||||||:::|||   |||:::::::||
   91 AlaSerAsnLeuGlyArgLeuGluAspTyrAsnMetAlaArgPheAsnLy 107

  498 ATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGG. 546
      ||||||||||:::::   |||:::      |||:::          |||
  107 sPheValThrGluValAlaProIleAlaProThrAspAlaGlyGlyGlyL 124

  547 ..CGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGG 594
            |||   |||:::|||::::::   :::   ||||||:::|||
  124 euAspThrTyrLysAspLysAsnArgPheSerSerPheValArgValGly 140

  595 GCAGGCAGGCAA......TATTGGCGATCTGATGAAGATGAGCCCAATAA 638
      ||||||||||||      |||   ::::::          :::   :::||
  141 AlaGlyArgGlnLeuValTyrGluLysGlyAlaTyrHisGlnGluGlyAs 157

  639 CCGCGAAAGTTCATAT.........CATATTGCAAGTGCGTATTCTTGGC 679
      |   |||::::::|||           :::::::::||||||   :::
  157 n...GluLysGlyTyrAspLeuArgAspLeuSerGlnAlaTyrArgTyrA 173

  680 TCGTTGGTGGC.................................AAT 693
        ::::::|||                                  |||
  173 laIleAlaGlyThrProTyrLysAspIleAsnIleAspGlnThrMetAsn 189

  694 ACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGA 743
      |||   :::      |||   |||:::   :::::::::   ::::::||
  190 ThrGluGlyLeuIleGlyPheGlyAsnHisAsnThrHisTyrSerAlaGl 206

  744 AAAAATTAAACATAGC...CCATATGGTTTTTTACCAACAGGAGGCTCAT 790
      |::::::|||:::::               |||   :::   |||   :
  206 uGluLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrGlyValL 223

  791 TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
      ::|||||||||||||||||:::|||   :::|||   ||||||::::::
  223 euGlyAspSerGlySerProLeuPheAlaPheAspLysGlnLysAsnGln 239

  841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAG 890
      |||::::::   |||:::   :::                |||   ||||||
  240 TrpValPheLeuGlyThrTyrAspTyrTrpAlaGlyTyr...GlyLys.. 254

  891 CAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTG 940
                          |||:::|||         :::|||:::
  255 .....................LysSerTrpGlnGluTrpAsnIleTyr. 263

  941 CTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATAC 990
                                                 |||
  264 ........................................LysLys 265

  991 TCTTTTAACGAC..............GATAATAATGGCACAGGAAAAAT 1025
      ::::|||   |||              ||||||   ||||||   |||
  266 GluPheAlaAspLysIleLysGlnArgAspAsnAlaGlyThrIleLysGl 282

 1026 CAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAA 1075
      |||:::::::|||:::   |||               :::         |
```

```
 282 yAsnGlyGluHisHisTrpAsn.............IleThrPheGlyT 294

1076 CCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTT 1125
     ||            |||      :::::::::||||||
 294 hr...........AsnSerHisIleGlySerThrAla........... 302

1126 TATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGG 1175
       :::   ||||||:::   ::::::              ||||||||
 303 ValArgLeuAlaGlyAsnGluArgAsp...........AlaAsnAsnGl 315

1176 AGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCA 1225
     |:::|||:::::::|||      |||   :::|||   |||:::|||   :
 315 yGlnAsnValThrPhe......GluAspAsnGlyThrLeuValLeuAspG 330

1226 GCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACG 1275
     ::|||||||||||||||||||||||||||:::|||:::|||||||:::|||
 330 lnAsnIleAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrThr 346

1276 GTCTCGCCTGAAAATAACGAA...ACTTGGCAAGGCGCGGGCGTTCATAT 1322
     |||:::      ||||||:::   ||||||   |||||||||:::   ::
 347 ValLysGlyAlaAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspVa 363

1323 CAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACC 1372
     :::::::   :::   |||:::|||:::|||::::::      :::|||||
 363 lAlaAspGlyLysLysValValTrpGlnValLysAsnProAsnGlyAspA 380

1373 GCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAA 1422
     |||||:::||||||||||||||||||||:::::::::::   |||
 380 rgLeuAlaLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyVal 396

1423 AACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCA 1472
     |||||||||:::::::::|||||||||||||||||||||||:::|||::
 397 AsnGlnGlyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnLy 413

1473 GGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCA 1522
     :||||||::::   |||   ||||||||||||:::::|||:::||||
 413 sAlaAspSerAsnGlnLysValGlnAlaPheSerGlnValGlyIleValS 430

1523 GCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGAC 1572
     ||||||||||||||:::   ||||||:::::::||||||:::||||||||
 430 erGlyArgGlyThrLeuValLeuAsnSerSerAsnGlnIleAsnProAsp 446

1573 AAACTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTC 1622
     :::|||||||||||||||||||||||||||||||   ||||||:::::
 447 AsnLeuTyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAs 463

1623 GCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCA 1672
     :|||:::|||:::::::|||:::|||:::|||||||||||||   |||||||
 463 pLeuThrPheGluHisIleArgAsnValAspGluGlyAlaArgIleValA 480

1673 ACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGAT 1722
     ||||||||                |||||||:::|||:::||||||   |||:::
 480 snHisAsnThrGlyHisAlaSerThrIleThrLeuThrGly...LysSer 495

1723 ATTGCTACA 1731
     :::   |||
 496 LeuIleThr 498

seq_name: sp_bacteria:Q51172

seq_documentation_block:
ID   Q51172      PRELIMINARY;      PRT;   496 AA.
AC   Q51172;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NGC80;
RX   MEDLINE=95302961; PubMed=7783620;
RA   Lomholt H., Poulsen K., Mogens K.;
RT   "Comparative characterization of the iga gene encoding IgA1 protease
RT   in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
RT   influenzae.";
RL   Mol. Microbiol. 15:495-506(1995).
DR   EMBL; X82479; CAA57862.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     496    496
SQ   SEQUENCE   496 AA;  54865 MW;  BEFC7A6003FE5D88 CRC64;


alignment_scores:
            Quality:  865.00              Length:    534
              Ratio:  2.437                Gaps:     20
 Percent Similarity:  66.479   Percent Identity: 39.326

alignment_block:
 US-09-303-518D-649 x Q51172   ..

 Align seg 1/1  to: Q51172  from: 1  to: 496

  202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
      |||   |||::::::|||   ||||||:::|||:::   :::|||:::::
    1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

  252 AATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAAC...G 298
      ::::::::::::|||||||||||||||||||:::::   |||
   17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnArgA 34

  299 GCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAAC 348
        :::   :::::::::      ||||||   |||||||||   |||
   34 rgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50

  349 GGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGAT.. 396
        |||   ::      :::|||   ::::::   :::   |||
   51 LysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAspVa 67

  397 ....CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAG 442
          :::   :::      |||:::   |||   :::|||::|||:::
   67 lSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAspTyrGluP 84

  443 CAGGGACTAAAGGCCATCCTTATGGCGGC.............GATTAT 477
        :::   :::   |||   :::|||:::                ||||||
   84 roAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuGluAspTyr 99

  478 CATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAAT 527
      :::|||   ||||||:::||||||||||::::::   |||:::
  100 AsnMetAlaArgLeuAsnLysPheValThrGluValAlaProIleAlaPr 116

  528 GACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAATAATTACC 574
       ||||||      |||   :::   |||   |||:::|||::::::
  116 oThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysAsnArgPheS 133

  575 CTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAA 624
        :::   ||||||:::|||||||||   |||:::
  133 erGluPheValArgValGlyAlaGlyThrGlnPhe.............. 144

  625 GATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTC 674
        |||   |||:::|||   :::      :::::::::   ||||||
  145 ...GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArgAlaTyrAr 160

  675 TTGGCTCGTTGGTGGCAATACCTTT.............GCACAAAATG 709
        :::   :::::::|||:::   :::                :::::::|||
  160 gTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValThrSerAsnL 177

  710 GATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGC 759
```

```
      :::      |||   ::::::::::|||::::::::::      ||||||
  177 euAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLysHisHisSer 193

  760 CCA.....TATGGTTTTTTACCAACAGGAGG 785
      |||                  :::   |||   :::  . ::
  194 ProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuThrAsnTyrAl 210

  786 CTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGC 835
      :   :::||||||||||||||||||||::|||   ||||||   |||:::: 
  210 aValLeuGlyAspSerGlySerProLeuPheAlaTyrAspLysGlnGluL 227

  836 AAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGA 885
      :::::|||::::::   |||:::   :::             |||
  227 ysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGlyTyr..... 241

  886 AAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG...TTCTATGATGA 932
      :::::::|||:::::::|||                :::|||   :::|||   ::
  242 GlnLysAsnSerTrpGln.....GluTrpAsnIleTyrLysLy 254

  933 AATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATG 982
      :   ||||||                          |||   :::|||:::
  254 sGluPheAlaAsp.....GluIleLysGlnArg. 263

  983 GGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAATCAATGCC 1032
                      ||||||   ||||||   |||   |||:::
  264 .....AspAsnAlaGlyThrIleLysGlyAsnGly 273

 1033 AAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCA 1082
      :::|||:::   ::::::                |||   |||
  274 GluHisHisTrpLysThr.....ThrGlyThr..... 282

 1083 ATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATG 1132
           |||      :::::::::|||                      |
  283 .....AsnSerHisIleGlySerThr.....A 290

 1133 CTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAAT 1182
      ||:::      :::|||:::   |||         ||||||||:::|||
  290 laValArgLeuAlaAsnAsnGluArgAsp...AlaAsnAsnGlyGlnAsn 305

 1183 ATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAACAT 1232
      ::::::|||      |||::::::|||   |||:::|||   :::||||| 
  306 ValThrPhe......GluAsnAsnGlyThrLeuValLeuAspGlnAsnIl 320

 1233 CAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCTCGC 1282
      |||||||||||||||||||||:::|||:::|||||:::|||||:::
  320 eAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrThrValLysG 337

 1283 CTGAAAATAACGAA...ACTTGGCAAGGCGCGGGCGTTCATATCAGTGAA 1329
           ||||||:::   ||||||   ||||||||||:::   :::::::::
  337 lyIleAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspValAlaAsp 353

 1330 GACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTC 1379
         :::   |||:::|||:::|||::::::      :::||||||||::
  354 GlyLysLysValValTrpGlnValLysAsnProAsnGlyAspArgLeuAl 370

 1380 CAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAG 1429
      :|||||||||||||||||||:::::::::::   |||   |||||||
  370 aLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyValAsnGlnG 387

 1430 GCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGAC 1479
      ||:::::::::|||||||||||||||||||||||:::|||||||||||
  387 lyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnGlnAlaAsp 403

 1480 GATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAG 1529
           |||   ||||||||||:::::::|||:::||||||||||
  404 AlaAspLysLysValGlnAlaPheSerGlnValGlyIleValSerGlyAr 420

 1530 GGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCT 1579
      |   |||:::   ||||||:::::|||||:::||||||||||:::||||
  420 gArgThrLeuValLeuAsnSerSerAsnGlnIleAsnProAspAsnLeuT 437

 1580 ATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCG 1629
      ||||||||||||||||||||||||||||   ||||||::::::|||:::

  437 yrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThr 453

 1630 TTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAA 1679
      |||:::::::|||:::|||:::||||||||||||   ||||||||||||
  454 PheGluHisIleArgAsnValAspGluGlyAlaArgIleValAsnHisAs 470

 1680 TCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTA 1729
      |           ||||||:::|||:::||||||   |||::::::   |
  470 nThrGlyHisAlaSerThrIleThrLeuThrGly...LysSerLeuIleT 486

 1730 CA 1731
      ||
  486 hr 486
```

seq_name: sp_bacteria:Q51168

seq_documentation_block:

```
ID   Q51168      PRELIMINARY;      PRT;   508 AA.
AC   Q51168;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NGC65;
RX   MEDLINE=95302961; PubMed=7783620;
RA   Lomholt H., Poulsen K., Mogens K.;
RT   "Comparative characterization of the iga gene encoding IgA1 protease
RT   in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
RT   influenzae.";
RL   Mol. Microbiol. 15:495-506(1995).
DR   EMBL; X82484; CAA57867.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     508    508
SQ   SEQUENCE   508 AA;  55927 MW;  23C65AFFBC493291 CRC64;
```

alignment_scores:

```
            Quality:  865.00              Length:    558
              Ratio:  2.403                 Gaps:     22
 Percent Similarity:  64.516   Percent Identity: 38.889
```

alignment_block:

```
 US-09-303-518D-649 x Q51168   ..

 Align seg 1/1  to: Q51168  from: 1  to: 508

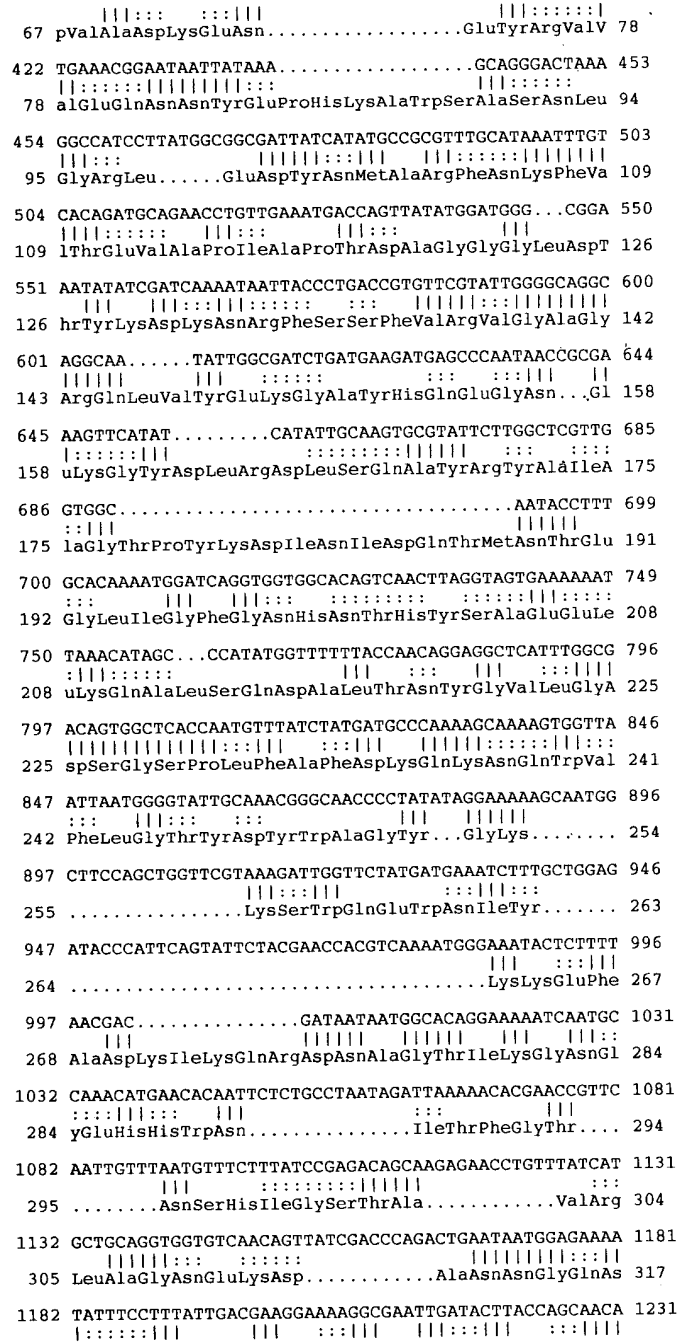
     202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
         |||   |||::::::|||   ||||||:::|||:::   :::|||:::::
       1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnIleGlyAsnAl 17

     252 AATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAAC...G 298
         :::::::::::::||||||||||||||||||||||:::::   |||
      17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnLysA 34

     299 GCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT... 345
           :::|||::::::|||      ||||||   |||||||||   |||
      34 rgIleAlaThrValValAspProGlnTyrAlaValSerValLysHisAla 50

     346 .....AACGGCGGCTATAACAACGTTGATTT 371
                                  |||   ||||||:::   :::
      51 LysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAsnAs 67

     372 TGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATTG 421
```



seq_name: sp_bacteria:Q51170

seq_documentation_block:

ID Q51170 PRELIMINARY; PRT; 496 AA.
AC Q51170;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE IGA1 PROTEASE (FRAGMENT).
GN IGA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HF54;
RX MEDLINE=95302961; PubMed=7783620;
RA Lomholt H., Poulsen K., Mogens K.;
RT "Comparative characterization of the iga gene encoding IgA1 protease
RT in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
RT influenzae.";
RL Mol. Microbiol. 15:495-506(1995).
DR EMBL; X82473; CAA57856.1; -.
DR MEROPS; S06.001; -.
DR InterPro; IPR000710; IgA_S6.
DR Pfam; PF02395; IGA1; 1.
DR PRINTS; PR00921; IGASERPTASE.

```
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     496    496
SQ   SEQUENCE   496 AA;  54733 MW;  4BFF91D0D54BE877 CRC64;


alignment_scores:
            Quality:  864.00              Length:    541
              Ratio:  2.407                Gaps:      19
 Percent Similarity:  66.359   Percent Identity: 38.632

alignment_block:
 US-09-303-518D-649 x Q51170   ..

 Align seg 1/1  to: Q51170  from: 1  to: 496

     202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
           |||   |||::::::::|||   |||||||:::|||:::   :::|||:::::
       1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

     252 AATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAAC...G 298
           :::::::::::::|||||||||||||||||||||||:::::   |||
      17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnArgA 34

     299 GCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAAC 348
           :::   :::::::::      ||||||   |||||||||   |||
      34 rgThrLeuThrValIleAspProGlnTyrAlaValSerValLysHisVal 50

     349 GGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGAT.. 396
           |||   :::      :::|||   ::::::   :::   |||
      51 LysGlyAspGluIleSerTyrTyrGlyHisHisAsnGlyHisLeuAspVa 67

     397 ....CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATAAAG 442
           :::   :::      |||:::   |||   :::|||:::|||:::
      67 lSerAsnAspGluAsnGluTyrArgSerValAlaGlnAsnAspTyrGluP 84

     443 CAGGGACTAAAGGCCATCCTTATGGCGGC.............GATTAT 477
           :::   :::   |||   :::|||:::              ||||||
      84 roAsnLysAsnTrpHis...HisGlyAsnGlnGlyArgLeuGluAspTyr 99

     478 CATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAAT 527
         :::|||   ||||||:::||||||||||||:::::   |||:::
     100 AsnMetAlaArgLeuAsnLysPheValThrGluValAlaProIleAlaPr 116

     528 GACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAATAATTACC 574
          ||||||        |||   :::   |||   |||:::|||::::::
     116 oThrSerAlaGlyGlyGlyValGluThrTyrLysAspLysAsnArgPheS 133

     575 CTGACCGTGTTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAA 624
           :::   ||||||:::|||||||||   |||:::
     133 erGluPheValArgValGlyAlaGlyThrGlnPhe.............. 144

     625 GATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTC 674
           |||   |||:::|||   :::      :::::::::   ||||||
     145 ...GluTyrAsnSerArgTyrAsnMetThrGluLeuSerArgAlaTyrAr 160

     675 TTGGCTCGTTGGTGGCAATACCTTT.............GCACAAAATG 709
           :::   ::::::|||:::   :::               :::::::|||
     160 gTyrAlaIleAlaGlyThrProTyrGlnAspValAsnValThrSerAsnL 177

     710 GATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGC 759
           :::      |||   :::::::::|||:::::::::      ||||||
     177 euAsnGlnGluGlyLeuIleGlyPheGlyAspAsnSerLysHisHisSer 193

     760 CCA.....................TATGGTTTTTTACCAACAGGAGG 785
         |||                          :::   |||   :::   ::
     194 ProGluLysLeuLysGluValLeuSerGlnAsnAlaLeuThrAsnTyrAl 210

     786 CTCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGC 835
         :   :::|||||||||||||||||||:::|||   ||||||   |||::::
     210 aValLeuGlyAspSerGlySerProLeuPheAlaTyrAspLysGlnGluL 227

     836 AAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGA 885
           :::::|||::::::   |||:::   :::            |||
     227 ysArgTrpValPheLeuGlyAlaTyrAspTyrTrpAlaGlyTyr..... 241

     886 AAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG...TTCTATGATGA 932
         :::::|||:::::|||           :::|||   :::|||   ::
     242 GlnLysAsnSerTrpGln...........GluTrpAsnIleTyrLysLy 254

     933 AATCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATG 982
         :   ||||||                      |||   :::|||:::
     254 sGluPheAlaAsp.....................GluIleLysGlnArg. 263

     983 GGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAATCAATGCC 1032
                               ||||||   ||||||   |||      :::
     264 ...................AspAsnAlaGlyThrIleLysGlyTyrGly 273

    1033 AAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCA 1082
         :::|||:::   ::::::                    |||   |||
     274 GluHisHisTrpLysThr.................ThrGlyThr..... 282

    1083 ATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATG 1132
                |||      :::::::::||||||   :::   :::   :::
     283 .......AsnSerHisIleGlySerThrAlaValArgLeuAlaGlyAsnG 297

    1133 CTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAAT 1182
               |||:::                          ||||||||||:::|||
     297 luArgGlyAla.....................AsnAsnGlyGlnAsn 305

    1183 ATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAACAT 1232
         :::::::|||      |||::::::|||   |||:::|||   :::||||||
     306 ValThrPhe......GluAsnAsnGlyThrLeuValLeuAspGlnAsnIl 320

    1233 CAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCTCGC 1282
         ||||||||||||||||||||||||:::|||:::||||||:::|||||:::
     320 eAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrThrValLysG 337

    1283 CTGAAAATAACGAA...ACTTGGCAAGGCGCGGGCGTTCATATCAGTGAA 1329
               ||||||      ||||||   |||||||||:::   :::::::::
     337 lyAlaAsnAsnGlyIleThrTrpLeuGlyAlaGlyIleAspValAlaAsp 353

    1330 GACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTC 1379
            :::   |||:::|||:::|||::::::      :::||||||||||::
     354 GlyLysLysValValTrpGlnValLysAsnProAsnGlyAspArgLeuAl 370

    1380 CAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAG 1429
         :|||||||||||||||||||||:::::::::::   |||   |||||||
     370 aLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyValAsnGlnG 387

    1430 GCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGAC 1479
         ||:::::::::|||||||||||||||||||||||:::||||||||||||
     387 lyGlnLeuLysValGlyAspGlyThrValIleLeuAsnGlnGlnAlaAsp 403

    1480 GATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAG 1529
                |||   |||||||||||||:::::|||:::||||||||||
     404 AlaAspLysLysValGlnAlaPheSerGlnValGlyIleValSerGlyAr 420

    1530 GGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCT 1579
         |   |||:::   ||||||:::   :::|||:::||||||:::::||||
     420 gArgThrLeuValLeuAsnSerProAspGlnIleAsnProAsnAsnLeuT 437

    1580 ATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCG 1629
         ||||||||||||||||||||||||||   ||||||:::::|||:::
     437 yrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThr 453

    1630 TTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAA 1679
         |||:::::::|||:::|||:::||||||||||||   ||||||||||||||
     454 PheGluHisIleArgAsnValAspGluGlyAlaArgIleValAsnHisAs 470

    1680 TCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTA 1729
         |   |||:::   ||||||:::|||:::||||||   |||::::::   |
     470 nThrAspArgAlaSerThrIleThrLeuThrGly...LysSerLeuIleT 486

    1730 CAACCGGCAATAACAACAGCTTG 1752
         ||:::   :::|||   |||:::
```

```
    486 hrAlaProGlnAsnLeuSerVal 493

seq_name: sp_bacteria:Q51164

seq_documentation_block:
ID   Q51164      PRELIMINARY;      PRT;   508 AA.
AC   Q51164;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SM1166;
RX   MEDLINE=95302961; PubMed=7783620;
RA   Lomholt H., Poulsen K., Mogens K.;
RT   "Comparative characterization of the iga gene encoding IgA1 protease
RT   in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
RT   influenzae.";
RL   Mol. Microbiol. 15:495-506(1995).
DR   EMBL; X82486; CAA57869.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     508    508
SQ   SEQUENCE   508 AA;  56025 MW;  FEAF91F9BE01B38F CRC64;


alignment_scores:
            Quality:  856.00            Length:    551
              Ratio:  2.404               Gaps:     22
 Percent Similarity:  64.610   Percent Identity: 38.657

alignment_block:
 US-09-303-518D-649 x Q51164   ..

 Align seg 1/1  to: Q51164  from: 1  to: 508

     202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
         |||   |||::::::|||   ||||||:::|||:::   :::|||:::::
       1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

     252 AATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAAC...G 298
         ::::::::::::|||||||||||||||||||:::::   |||
      17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnLysA 34

     299 GCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT... 345
           :::|||::::::|||      ||||||   |||||||||   |||
      34 rgIleAlaThrValValAspProGlnTyrAlaValSerValLysHisAla 50

     346 .......................AACGGCGGCTATAACAACGTTGATTT 371
                                   |||   |||||||:::   :::
      51 LysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAsnAs 67

     372 TGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATTG 421
         |||:::   :::|||                        |||::::::|
      67 pValAlaAspLysGluAsn.................GluTyrArgValV 78

     422 TGAAACGGAATAATTATAAA.................GCAGGGACTAAA 453
         ||::::::|||||||||:::                  |||::::::
      78 alGluGlnAsnAsnTyrGluProHisLysAlaTrpSerAlaSerAsnLeu 94

     454 GGCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAATTTGT 503
         |||:::            ||||||:::|||   |||:::::|||||||
      95 GlyArgLeu......GluAspTyrAsnMetAlaArgPheAsnLysPheVa 109

     504 CACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGG...CGGA 550
         ||||::::::   |||:::      |||:::         |||
     109 lThrGluValAlaProIleAlaProThrAspAlaGlyGlyGlyLeuAspT 126

     551 AATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGC 600
           |||   |||:::|||::::::   :::   ||||||:::|||||||||
     126 hrTyrLysAspLysAsnArgPheSerSerPheValArgValGlyAlaGly 142

     601 AGGCAA......TATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGA 644
         ||||||      |||   ::::::         :::   :::|||   ||
     143 ArgGlnLeuValTyrGluLysGlyAlaTyrHisGlnGluGlyAsn...Gl 158

     645 AAGTTCATAT.........CATATTGCAAGTGCGTATTCTTGGCTCGTTG 685
         |:::::|||           :::::::::||||||   :::   ::::
     158 uLysGlyTyrAspLeuArgAspLeuSerGlnAlaTyrArgTyrAlaIleA 175

     686 GTGGC.................................AATACCTTT 699
         ::|||                                 ||||||
     175 laGlyThrProTyrLysAspIleAsnIleAspGlnThrMetAsnThrGlu 191

     700 GCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAAT 749
         :::      |||   |||:::   :::::::::   ::::::|||:::::
     192 GlyLeuIleGlyPheGlyAsnHisAsnThrHisTyrSerAlaGluGluLe 208

     750 TAAACATAGC...CCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCG 796
         :|||::::::          |||   :::   |||   :::||||
     208 uLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrGlyValLeuGlyA 225

     797 ACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTA 846
         |||||||||||||:::|||   :::|||   ||||||::::::|||:::
     225 spSerGlySerProLeuPheAlaPheAspLysGlnLysAsnGlnTrpVal 241

     847 ATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAATGG 896
         :::   |||:::   :::            |||   ||||||
     242 PheLeuGlyThrTyrAspTyrTrpAlaGlyTyr...GlyLys........ 254

     897 CTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTGGAG 946
                         |||:::|||         :::|||:::
     255 ..............LysSerTrpGlnGluTrpAsnIleTyr....... 263

     947 ATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCTTTT 996
                                                 |||   :::|||
     264 ....................................LysLysGluPhe 267

     997 AACGAC..............GATAATAATGGCACAGGAAAAATCAATGC 1031
           |||                ||||||   ||||||   |||   |||::
     268 AlaAspLysIleLysGlnArgAspAsnAlaGlyThrIleLysGlyAsnGl 284

    1032 CAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTC 1081
         :::|||:::   |||              :::                |||
     284 yGluHisHisTrpAsn..............IleThrPheGlyThr.... 294

    1082 AATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCAT 1131
                |||      :::::::::||||||              :::
     295 ........AsnSerHisIleGlySerThrAla............ValArg 304

    1132 GCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAA 1181
         ||||||:::   ::::::                |||||||||:::||
     305 LeuAlaGlyAsnGluArgAsp............AlaAsnAsnGlyGlnAs 317

    1182 TATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAACA 1231
         |:::::|||      |||   :::|||   |||:::|||   :::||||
     317 nValThrPhe......GluAspAsnGlyThrLeuValLeuAspGlnAsnI 332

    1232 TCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCTCG 1281
         |||||||||||||||||||||||:::|||:::||||||:::||||||:::
     332 leAsnGlnGlyAlaGlyGlyLeuPhePheLysGlyAspTyrThrValLys 348

    1282 CCTGAAAATAACGAA...ACTTGGCAAGGCGCGGGCGTTCATATCAGTGA 1328
               ||||||:::   ||||||   |||||||||:::   ::::::::
     349 GlyAlaAsnAsnAspIleThrTrpLeuGlyAlaGlyIleAspValAlaAs 365

    1329 AGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGT 1378
         :   :::   |||:::|||:::|||::::::      :::||||||||:
```

```
 365 pGlyLysLysValValTrpGlnValLysAsnProAsnGlyAspArgLeuA 382

1379 CCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAA 1428
     ::|||||||||||||||||||||||:::::::::::::   |||   ||||||
 382 laLysIleGlyLysGlyThrLeuGluIleAsnGlyThrGlyValAsnGln 398

1429 GGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGA 1478
     |||:::::::::|||||||||||||||||||||||||:::|||:::|||||
 399 GlyGluLeuLysValGlyAspGlyThrValIleLeuAsnGlnLysAlaAs 415

1479 CGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCA 1528
     |:::::   |||   :::|||||||||:::::|||:::|||   ||||
 415 pSerAsnGlnLysValSerAlaPheSerGlnValGlyIleValArgGlyA 432

1529 GGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTC 1578
     |||||||||:::   ||||||:::   :::|||:::||||||:::::|||
 432 rgGlyThrLeuValLeuAsnSerProAspGlnIleAsnProAsnAsnLeu 448

1579 TATTTCGGCTTTCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTC 1628
     |||||||||||||||||||||||||||||   ||||||:::::|||::
 449 TyrPheGlyPheArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuTh 465

1629 GTTCCACCGTATTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACA 1678
     :|||:::::::|||:::|||:::||||||||||||   ||||||||||||
 465 rPheGluHisIleArgAsnValAspGluGlyAlaArgIleValAsnHisA 482

1679 ATCAAGACAAAGAATCCACCGTTACCATTACAGGCAATAAAGATATTGCT 1728
     ||   |||:::   ||||||:::|||:::||||||   |||:::::::
 482 snThrAspArgAlaSerThrIleThrLeuThrGly...LysSerLeuIle 497

1729 ACA 1731
     |||
 498 Thr 498
```

seq_name: sp_bacteria:Q51173

seq_documentation_block:

```
ID   Q51173      PRELIMINARY;      PRT;   507 AA.
AC   Q51173;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Neisseria meningitidis.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=487;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NGO93;
RX   MEDLINE=95302961; PubMed=7783620;
RA   Lomholt H., Poulsen K., Mogens K.;
RT   "Comparative characterization of the iga gene encoding IgA1 protease
RT   in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
RT   influenzae.";
RL   Mol. Microbiol. 15:495-506(1995).
DR   EMBL; X82482; CAA57865.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     507    507
SQ   SEQUENCE   507 AA;  55956 MW;  80E7BA3F04EBEA82 CRC64;
```

alignment_scores:

```
            Quality:  850.50              Length:    547
              Ratio:  2.437                 Gaps:     19
 Percent Similarity:  63.803   Percent Identity: 38.757
```

alignment_block:

US-09-303-518D-649 x Q51173 ..

Align seg 1/1 to: Q51173 from: 1 to: 507

```
 202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
     |||   |||:::::|||   ||||||:::|||:::   :::|||:::::
   1 AlaThrAspLeuSerValLysAsnLysGlnGlyGlnAsnIleGlyAsnAl 17

 252 AATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAAC...G 298
     :::::::::::::|||||||||||||||||||||:::::   |||
  17 aLeuSerAsnValProMetIleAspPheSerValAlaAspValAsnLysA 34

 299 GCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT... 345
     :::|||:::::|||      ||||||   |||||||||   |||
  34 rgIleAlaThrValValAspProGlnTyrAlaValSerValLysHisAla 50

 346 ......................AACGGCGGCTATAACAACGTTGATTT 371
                                |||   |||||:::   :::
  51 LysAlaGluValHisThrPheTyrTyrGlyGlnTyrAsnGlyHisAsnAs 67

 372 TGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATTG 421
         |||:::   :::|||                  |||:::::::|
  67 pValAlaAspLysGluAsn.................GluTyrArgValV 78

 422 TGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGC 471
     ||:::::|||||||||:::                 |||         |||
  78 alGluGlnAsnAsnTyrGluPro.............HisLysAlaTrpGly 90

 472 ......................GATTATCATATGCCGCGTTTGCATAA 497
                                |||||||:::|||   |||:::::::||
  91 AlaSerAsnLeuGlyArgLeuGluAspTyrAsnMetAlaArgPheAsnLy 107

 498 ATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGG. 546
     ||||||||||::::::   |||:::      |||:::         |||
 107 sPheValThrGluValAlaProIleAlaProThrAspAlaGlyGlyGlyL 124

 547 ..CGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGG 594
           |||   |||:::|||::::::   :::   ||||||:::|||
 124 euAspThrTyrLysAspLysAsnArgPheSerSerPheValArgValGly 140

 595 GCAGGCAGGCAA...TATTGGCGATCTGATGAAGATGAGCCCAATAACCG 641
     ||||||||||||      :::::::::         :::   ::::::::
 141 AlaGlyArgGlnLeuValTyrGluLysGlyAlaTyrHisGlnGluGlyLy 157

 642 CGAAAGTTCATAT.........CATATTGCAAGTGCGTATTCTTGGCTCG 682
     :|||:::::::|||           :::::::::||||||   :::   :
 157 sGluLysGlyTyrAspLeuArgAspLeuSerGlnAlaTyrArgTyrAlaI 174

 683 TTGGTGGC.....................................AATACC 696
     ::::::|||                                      ||||||
 174 leAlaGlyThrProTyrLysAspIleAsnIleAspGlnThrMetAsnThr 190

 697 TTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAA 746
     :::      |||   |||:::   :::   :::   ::::::|||::
 191 GluGlyLeuIleGlyPheGlyAsnHisAsnLysGlnTyrSerAlaGluGl 207

 747 AATTAAACATAGC...CCATATGGTTTTTTACCAACAGGAGGCTCATTTG 793
     ::::|||::::::               |||   :::   |||   :::|
 207 uLeuLysGlnAlaLeuSerGlnAspAlaLeuThrAsnTyrGlyValLeuG 224

 794 GCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGG 843
     |||||||||||||||||:::|||   :::|||   ||||||:::::::|||
 224 lyAspSerGlySerProLeuPheAlaPheAspLysGlnLysAsnGlnTrp 240

 844 TTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAAAAAGCAA 893
     ::::::   |||:::   :::            |||      ||||||
 241 ValPheLeuGlyThrTyrAspTyrTrpAlaGlyTyrGlyLysLysSerTr 257

 894 TGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATCTTTGCTG 943
                 :::   |||      |||   |||||||||
 257 pGlnGluTrpAsnIleTyrLysLysGluPheAlaAspGluIle..... 271

 944 GAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAAATACTCT 993
                                   :::|||:::
```

```
 272 .............................LysGlnArg............ 274

 994 TTTAACGACGATAATAATGGCACAGGAAAAATCAATGCCAAACATGAACA 1043
                 ||||||   ||||||   |||      :::::|||:::
 275 .........AspAsnAlaGlyThrIleLysGlyTyrGlyGluHisHisTr 288

1044 CAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAATTGTTTAATG 1093
       ::::::                  |||   |||            |||
 288 pLysThr..................ThrGlyThr............AsnS 295

1094 TTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGT 1143
        ::::::::::|||                              |||:::
 295 erHisIleGlySerThr....................AlaValArgLeu 304

1144 GTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTAT 1193
     :::|||:::   |||         ||||||||||:::|||:::::|||
 305 AlaAsnAsnGluArgAsp...AlaAsnAsnGlyGlnAsnValThrPhe.. 319

1194 TGACGAAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGTG 1243
         |||::::::|||   |||:::|||   :::|||||||||||||||||
 320 ....GluAsnAsnGlyThrLeuValLeuAspGlnAsnIleAsnGlnGlyA 335

1244 CTGGAGGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAAC 1293
     ||||||||||:::|||:::|||||:::||||||:::      ||||||
 335 laGlyGlyLeuPhePheLysGlyAspTyrThrValLysGlyAlaAsnAsn 351

1294 GAA...ACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGT 1340
     :::   ||||||   ||||||||||:::   :::::::::   :::   ||
 352 AspIleThrTrpLeuGlyAlaGlyIleAspValAlaAspGlyLysLysVa 368

1341 TACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCA 1390
     |:::|||:::|||::::::      :::||||||||||:::||||||||||
 368 lValTrpGlnValLysAsnProAsnGlyAspArgLeuAlaLysIleGlyL 385

1391 AAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAGC 1440
     ||||||||||:::::::::::   |||   |||||||||:::::::::
 385 ysGlyThrLeuGluIleAsnGlyThrGlyValAsnGlnGlyGlnLeuLys 401

1441 GTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAA 1490
     ||||||||||||||||||||||:::|||:::||||||       ||
 402 ValGlyAspGlyThrValIleLeuAsnGlnLysAlaAspAlaAspLysLy 418

1491 AAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGC 1540
     |   ||||||||||||:::::|||:::||||||||||||   |||:::
 418 sValGlnAlaPheSerGlnValGlyIleValSerGlyArgArgThrLeuV 435

1541 AACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTTT 1590
       ||||||::::::||||||:::||||||||||:::|||||||||||||||
 435 alLeuAsnSerSerAsnGlnIleAsnProAspAsnLeuTyrPheGlyPhe 451

1591 CGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTAT 1640
     ||||||||||||||||   ||||||::::::|||:::|||::::::||
 452 ArgGlyGlyArgLeuAspAlaAsnGlyAsnAspLeuThrPheGluHisIl 468

1641 TCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAG 1690
     |:::|||:::||||||||||||   ||||||||||||||
 468 eArgAsnValAspGluGlyAlaArgIleValAsnHisAsnThrGlyHisA 485

1691 AATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACA 1731
       ||||||:::|||:::||||||   |||::::::   |||
 485 laSerThrIleThrLeuThrGly...LysSerLeuIleThr 497

seq_name: sp_bacteria:O69746

seq_documentation_block:
ID   O69746      PRELIMINARY;      PRT;   192 AA.
AC   O69746;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA PROTEASE (FRAGMENT).
OS   Neisseria lactamica.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=486;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NL3327;
RA   Sayers J.R., Vitowski S., Read R.;
RT   "Strains of Neisseria lactamica carry an IgA1 protease gene homologue
RT   and some express IgA1 protease activity.";
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ001740; CAA04965.1; -.
DR   MEROPS; S06.001; -.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     192    192
SQ   SEQUENCE   192 AA;  21385 MW;  E4C6FE902FFA08F8 CRC64;


alignment_scores:
            Quality:  717.50              Length:    200
              Ratio:  4.246                 Gaps:      3
 Percent Similarity:  84.500   Percent Identity: 70.000

alignment_block:
 US-09-303-518D-649 x O69746   ..

 Align seg 1/1  to: O69746  from: 1  to: 192

     184 GGCAAGTTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGG 233
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 GlyLysPheAlaValGlyAlaLysAspIleGluValTyrAsnLysLysGl 17

     234 GGAGTTGGTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTG 283
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 yGluLeuValGlyLysSerMetThrLysAlaProMetIleAspPheSerV 34

     284 TGGTGTCGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATATATTGTG 333
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 alValSerArgAsnGlyValAlaAlaLeuValGlyAspGlnTyrIleVal 50

     334 AGCGTGGCACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGG 383
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      51 SerValAlaHisAsnGlyGlyTyrAsnAsnValAspPheGlyAlaGluGl 67

     384 AAGAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATA 433
         |   |||||||||||||||||||:::|||:::|||||||||||||||||
      67 ySerAsnProAspGlnHisArgPheSerTyrGlnIleValLysArgAsnA 84

     434 ATTATAAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATG 483
         ||||||||         ::::::||||||||||:::|||||||||||||||
      84 snTyrLysPro...AspAsnSerHisProTyrAsnGlyAspTyrHisMet 99

     484 CCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAG 533
         ||||||||||||||||||||||||||||||||||||||||||||||||::
     100 ProArgLeuHisLysPheValThrAspAlaGluProValGluMetThrGl 116

     534 TTATATGGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTG 583
         :   |||   |||:::   |||   |||:::::::::|||||||:::||||
     116 yAspMetArgGlyAsnThrTyrSerAspLysGluLysTyrProGluArgV 133

     584 TTCGTATTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAGCCC 633
         ||||||||||:::|||:::::::|||||||||   |||:::|||:::
     133 alArgIleGlySerGlyHisHisTyrTrpArgTyrAspAspAspLys... 148

     634 AATAACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTCT......TG 677
                                 |||         :::||||||      ||
     149 ....................HisGlyAspLeuSerTyrSerGlyAlaTr 158

     678 GCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAG 727
         ||||:::|||||||||||||      |||:::  :::::::::|||:::|
     158 pLeuIleGlyGlyAsnThrHisMetGlnGlyTrpGlyAsnAsnGlyValV 175

     728 TCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCA 777
         ||:::|||:::::::::            :::   ||||||   :::|||
     175 alSerLeuSerGlyAspValArgHisAlaAsnAspTyrGlyProMetPro 191
```

```
seq_name: sp_bacteria:O69747

seq_documentation_block:
ID   O69747      PRELIMINARY;      PRT;   194 AA.
AC   O69747;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
OS   Neisseria lactamica.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=486;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NL3354;
RA   Sayers J.R., Vitowski S., Read R.;
RT   "Strains of Neisseria lactamica carry an IgA1 protease gene homologue
RT   and some express IgA1 protease activity.";
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ001739; CAA04964.1; -.
DR   MEROPS; S06.001; -.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     194    194
SQ   SEQUENCE   194 AA;  21514 MW;  D474FB2FC82DA988 CRC64;


alignment_scores:
           Quality:  714.00            Length:    200
             Ratio:  4.127               Gaps:      3
 Percent Similarity:  86.500   Percent Identity: 69.000

alignment_block:
 US-09-303-518D-649 x O69747   ..

 Align seg 1/1  to: O69747  from: 1  to: 194

     190 TTTGCAGTCGGGGCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTT 239
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 PheAlaValGlyAlaLysAspIleGluValTyrAsnLysLysGlyGluLe 17

     240 GGTCGGCAAATCAATGACAAAAGCCCCGATGATTGATTTTTCTGTGGTGT 289
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 uValGlyLysSerMetThrLysAlaProMetIleAspPheSerValValS 34

     290 CGCGTAACGGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTG 339
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      34 erArgAsnGlyValAlaAlaLeuValGlyAspGlnTyrIleValSerVal 50

     340 GCACATAACGGCGGCTATAACAACGTTGATTTTGGTGCGGAAGGAAGAAA 389
         |||||||||||||||||||||||||||||||||||||||||||||||  ||
      51 AlaHisAsnGlyGlyTyrAsnAsnValAspPheGlyAlaGluGlySerAs 67

     390 TCCCGATCAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTATA 439
         ||||||||||||||||||:::|||:::||||||||||||||||||||||
      67 nProAspGlnHisArgPheSerTyrGlnIleValLysArgAsnAsnTyrL 84

     440 AAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCGT 489
         ||        ::::::||||||||||:::|||||||||||||||||||||
      84 ysPro...AspAsnSerHisProTyrAsnGlyAspTyrHisMetProArg 99

     490 TTGCATAAATTTGTCACAGATGCAGAACCTGTTGAAATGACCAGTTATAT 539
         ||||||||||||||||||||||||||||||||||||||||||::::  ||
     100 LeuHisLysPheValThrAspAlaGluProValGluMetThrGlyAspMe 116

     540 GGATGGGCGGAAATATATCGATCAAAATAATTACCCTGACCGTGTTCGTA 589
         |   |||:::   |||   |||::::::::::||||||:::||||||||||
     116 tArgGlyAsnThrTyrSerAspLysGluLysTyrProGluArgValArgI 133

     590 TTGGGGCAGGCAGGCAATATTGGCGATCTGATGAAGATGAGCCCAATAAC 639
         |||||:::|||:::::::||||||||   |||:::|||:::   ::::::
     133 leGlySerGlyHisHisTyrTrpArgTyrAspAspAspAspLysHisGly 149

     640 CGCGAAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGG 689
         :::   ||||||           :::   :::||||||:::|||||
     150 ...AspLeuSerTyr...........SerGlyAlaTrpLeuIleGlyGl 161

     690 CAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTA 739
         |||||||      |||:::   :::::::::|||:::|||:::|||:::
     161 yAsnThrHisMetGlnGlyTrpGlyAsnAsnGlyValValSerLeuSerG 178

     740 GTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGCTCA 789
         :::::           :::   ||||||   :::|||   :::|||:::
     178 lyAspValArgHisAlaAsnAspTyrGlyProMetProIleAlaGlyAla 194

seq_name: sp_bacteria:Q48205

seq_documentation_block:
ID   Q48205      PRELIMINARY;      PRT;   462 AA.
AC   Q48205;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HK635;
RX   MEDLINE=95302961; PubMed=7783620;
RA   Lomholt H., Poulsen K., Mogens K.;
RT   "Comparative characterization of the iga gene encoding IgA1 protease
RT   in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
RT   influenzae.";
RL   Mol. Microbiol. 15:495-506(1995).
DR   EMBL; X82488; CAA57871.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
KW   Protease.
FT   NON_TER       1      1
SQ   SEQUENCE   462 AA;  51190 MW;  46B634CFF3EE7202 CRC64;


alignment_scores:
           Quality:  712.50            Length:    506
             Ratio:  2.213               Gaps:     21
 Percent Similarity:  63.636   Percent Identity: 36.166

alignment_block:
 US-09-303-518D-649 x Q48205   ..

 Align seg 1/1  to: Q48205  from: 1  to: 462

     202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
         |||   ::::::||||||   :::|||:::::::::   :::|||:::::
       1 AlaThrAsnValGluValArgAspLysAsnAsnHisSerLeuGlyAsnAl 17

     252 AATGACAAAAGCC...CCGATGATTGATTTTTCTGTGGTGTCGCGTAAC. 297
         ::::   ::::::   |||||||||||||||||||||||:::   |||
      17 aLeuProAsnGlyIleProMetIleAspPheSerValValAspValAsnL 34

     298 ..GGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT 345
          :::::::::||||||      ||||||||||||||||   |||
      34 ysArgIleGlyThrLeuValAspProGlnTyrIleValSerValLysHis 50

     346 AACGGCGGCTATAACAACGTTGATTTT.................GGTGC 377
             |||   |||   ||||||                     |||
      51 AlaHisGlnTyrMetAsn...AspPheTyrPheGlyHisTyrAsnGlyHi 66

     378 GGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAATTGTGAAAC 427
         :::       :::   |||::::::::::      |||::::::|||   :
```

```
 66 sArgAspValSerAspAspGluAsnLys......TyrSerValValThrG 81

428 GGAATAATTATAAAGCAGGGACTAAAGGCCAT......CCTTATGGCGGC 471
    ::|||||   :::   :::   :::   |||
 81 lnAsnAsnValAsnProAsnGluAsnTrpHisValAspLysArgLeuAsp 97

472 GATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGCAGAACCTGT 521
    ||||||:::|||||||||||||:::|||||||||||||::::::   |||::
 98 AspTyrAsnMetProArgLeuAsnLysPheValThrGluValAlaProTh 114

522 TGAAATGACCAGTTATATGGATGGG...CGGAAATATATCGATCAAAATA 568
    :   |||         |||      :::   |||   |||:::::::
114 rThrProThrLeuAlaGlyAspAspLeuGluThrTyrLysAspLysGluL 131

569 ATTACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAA......TATTGG 612
    ::|||   :::   ||||||:::|||||||||||||||      |||   .
131 ysTyrLeuSerPheValArgValGlyAlaGlyArgGlnLeuValTyrGlu 147

613 CGATCTGATGAA......GATGAGCCCAATAACCGCGAAAGTTCATATCA 656
    ::::::::::::      ::::::   ::::::   |||:::
148 LysGlySerHisHisValGluAspLysGluHisGlyGluAspLeuLysAs 164

657 TATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCA.... 702
     :::::::::||||||   :::   :::||||||:::   :::
164 pLeuSerAlaAlaTyrArgTyrAlaIleGlyGlyThrProTyrLysGlyI 181

703 ..........CAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGT 741
               :::   |||      |||   ::::::::::|||:::
181 leAsnIleAspProSerGlnSerLysLysGlyLeuIleGlyPheGlyAsp 197

742 GAAAAAATTAAACAT................................ 756
    ::::::      |||
198 SerArgGluAspHisValIleAsnSerLysThrLeuLeuSerGlnAspPr 214

757 ....AGCCCATATGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTG 802
        :::   ||||||   |||                ||||||||||
214 oLeuThrAsnTyrGlyValLeu.................GlyAspSerG 225

803 GCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAAT 852
    ||||||||:::|||   :::|||   |||::::::||||||::::::
225 lySerProLeuPheAlaPheAspLysGlnGlnAsnLysTrpValPheIle 241

853 GGGGTATTGCAAACGGGCAACCCCTAT.............ATAGGAAA 887
    |||                  ||||||                |||||
242 Gly.................ProTyrThrTyrTrpAlaGlyTyrGlyLy 252

888 AAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGG...TTCTATGATGAAA 934
    |::::::                  :::::::|||   :::|||   :::
252 sLysSerTrp.............GlnGluTrpAsnIleTyrLysSerG 264

935 TCTTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGG 984
      .|||:::   |||:::
264 lnPheThrLysAspVal............................... 269

985 AAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAATCAATGCCAA 1034
          :::|||   |||:::   |||      |||   |||:::::
270 .........LeuAsnLysAspSerAlaGlyLeuLeuLysGlyAsnThrGl 283

1035 ACATGAACACAATTCTCTGCCTAAT...AGATTAAAAACACGAACCGTTC 1081
     :::::::   :::|||      |||      :::   ::::::      :
283 nTyrAsnTrpThrSerAsnGlyAsnThrSerMetIleSerAsnGlySerG 300

1082 AATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCAT 1131
     ::|||::::::|||:::|||   ::::::::::::
300 luLeuLeuGluValAsnLeuPheAspAsnSerLys.............. 311

1132 GCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAA 1181
                  ::::::   |||   :::   |||   |||:::::
312 ..............HisThrAsnArgGluLysAlaAsnTyrGlyLysSe 323

1182 TATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAACA 1231
     :::::::|||      :::|||:::|||   |||   |||   ::::::|
323 rValThrPhe......GlnGlyAsnGlyThrLeuThrLeuLysAsnSerI 338

1232 TCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACGGTC... 1278
     ||||||||||||||||||||||:::|||:::|||:::::::||||||
 338 leAsnGlnGlyAlaGlyGlyLeuPhePheGluGlyAsnTyrThrValGlu 354

1279 TCGCCTGAAAATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCAGTGA 1328
     :::   ::::::|||   :::|||:::|||||||||:::   ::::::||
 355 GlySerSerAspAsnIleValTrpAsnGlyAlaGlyIleSerValAlaGl 371

1329 AGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGT 1378
     |   :::|||||||||||||||||||:::::      :::||||||||:
 371 uGlyLysThrValThrTrpLysValHisAsnProGlnSerAspArgLeuA 388

1379 CCAAAATCGGCAAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAA 1428
     ::|||||||||||||||||||   |||::::::|||||||||||||:::
 388 laLysIleGlyLysGlyThrLeuIleValGluGlyLysGlyGluAsnLys 404

1429 GGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGA 1478
     ||||||::::::|||||||||||||||||||||||||   ||||||||||
 405 GlySerLeuLysValGlyAspGlyThrValIleLeuLysGlnGlnAlaAs 421

1479 CGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCA 1528
     |   :::::|||   :::|||||||||:::::|||:::||||||||||
 421 pAlaAsnAsnLysValLysAlaPheSerGlnValGlyIleValSerGlyA 438

1529 GGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTC 1578
     ||:::||||||   ||||||   |||:::|||   :::|||::::::::::
 438 rgSerThrValValLeuAsnAspAspLysGlnValAspProAsnSerIle 454

1579 TATTTCGGCTTTCGCGGC 1596
     ||||||||||||||||||
 455 TyrPheGlyPheArgGly 460
```

seq_name: sp_bacteria:Q48204

seq_documentation_block:

```
ID   Q48204      PRELIMINARY;      PRT;   471 AA.
AC   Q48204;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=HK284;
RX   MEDLINE=95302961; PubMed=7783620;
RA   Lomholt H., Poulsen K., Mogens K.;
RT   "Comparative characterization of the iga gene encoding IgA1 protease
RT   in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
RT   influenzae.";
RL   Mol. Microbiol. 15:495-506(1995).
DR   EMBL; X82487; CAA57870.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     471    471
SQ   SEQUENCE   471 AA;  52559 MW;  68B57D31F21E6561 CRC64;
```

alignment_scores:

```
            Quality:  677.50              Length:     520
              Ratio:  2.111                Gaps:      24
 Percent Similarity:  61.731   Percent Identity: 35.000
```

alignment_block:

```
 US-09-303-518D-649 x Q48204   ..
```

Align seg 1/1 to: Q48204 from: 1 to: 471

```
202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
    |||   ::::::||||||   :::|||:::::::::   :::|||:::::
  1 AlaThrAsnValGluValArgAspLysAsnAsnGlnSerLeuGlySerAl 17

252 AATGACAAAAGCC...CCGATGATTGATTTTTCTGTGGTGTCGCGTAAC. 297
    ::::   ::::::   |||||||||||||||||||||||||:::   |||
 17 aLeuProAsnGlyIleProMetIleAspPheSerValValAspValAsnL 34

298 ..GGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT 345
        :::::::::||||||      ||||||||||||||||||   |||
 34 ysArgIleGlyThrLeuValAspProGlnTyrIleValSerValLysHis 50

346 AAC.........................GGCGGCTATAACAACGTTGA 368
                                |||   ||||||
 51 AlaHisLysHisIleAsnAspPheTyrPheGlyHisTyrAsn..... 64

369 TTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTACTTATAAAA 418
        |||   :::      :::   |||:::::::::      |||::::
 65 ....GlyHisArgAspValSerAspAspGluAsnLys......TyrSerV 78

419 TTGTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCAT......CCT 462
    ::|||   :::||||||   |||   :::         |||
 78 alValThrGlnAsnAsnValLysProAsnGluAspTrpHisValAspLys 94

463 TATGGCGGCGATTATCATATGCCGCGTTTGCATAAATTTGTCACAGATGC 512
            ||||||:::||||||||||||:::||||||||||||:::::
 95 ArgLeuAspAspTyrAsnMetProArgLeuAsnLysPheValThrGluVa 111

513 AGAACCTGTTGAAATGACCAGTTATATGGATGGG...CGGAAATATATCG 559
    :   |||:::      |||         |||      ::   |||. |
111 lAlaProThrThrProThrLeuAlaGlyAspAspLeuGluThrTyrLysA 128

560 ATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGCAGGCAATAT 609
    ||:::::::::|||||:::   ||||||:::|||||||||   |||:::
128 spLysGluLysTyrProSerPheValArgValGlyAlaGlyThrGlnPhe 144

610 TGGCGATCTGAT..............GAAGATGAGCCCAATAACCGCGA 644
        :::                  |||         :::|||:::::
145 ValTyrGluLysGlySerTyrTyrValGluLysThrThrArgAsnAsnAs 161

645 AAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATA 694
    :   ::::::   :::   :::||||||   :::   :::|||||||:::
161 pIleLysPhe...LeuAspGluAlaTyrArgTyrAlaIleGlyGlyThrP 177

695 CCTTTGCA..............CAAAATGGATCAGGTGGTGGCACAGTC 729
      :::                ::::   |||      |||   :::
177 roTyrGluGlyIleAsnIleAspProSerGlnSerLysLysGlyLeuIle 193

730 AACTTAGGTAGTGAAAAAATTAAACAT..................... 756
    ::::::|||::::::::::   :::|||
194 GlyPheGlyAspSerArgGluAsnHisValIleAspAlaLysThrLeuLe 210

757 ..............AGCCCATATGGTTTTTTACCAACAGGAGGCTCAT 790
                ::::   ||||||   |||
210 uSerGlnAspProLeuThrAsnTyrGlyValLeu.............. 221

791 TTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCAAAAG 840
      |||||||||||||||||||:::|||   :::|||   |||:::::::|||
222 ..GlyAspSerGlySerProLeuPheAlaPheAspLysGlnGlnAsnLys 237

841 TGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTAT.......... 879
    |||::::::   |||                  ||||||
238 TrpValPheIleGly.................ProTyrThrTyrTrpAl 248

880 ....ATAGGAAAAAGCAAT......GGCTTCCAGCTGGTTCGTAAAGATT 919
          ||||||::::::         :::::::::   :::|||||||
248 aGlyTyrGlyLysLysSerTrpGlnGluTrpAsnIleTyrLysLysAsp. 264

920 GGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCTACGAA 969
      |||   |||:::|||         |||:::   ::::::
```

```
265 ..PheAlaAspAsnIleLysLysArgAspAsnAlaGluAlaVal...... 278

970 CCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGG 1019
    |||   :::::::::::::|||   :::::::::   :::|||      ::
279 ProPheSerThrSerGluTyrHisTrpThrAsnThrThrAsnHisGlnSe 295

1020 AAAAATC...AATGCCAAACATGAACACAAT............TCTCTGC 1054
     ::::|||   |||:::   |||      :::            |||   |
295 rGluIleLysAsnThrAspHisThrIleThrValThrLeuProSerAspP 312

1055 CTAATAGATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCC 1104
     ||||||||||
312 roAsnArgLeu..................................... 315

1105 GAGACAGCAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAAC...AG 1151
                                                |||||   ::
316 .....................................ValAsnPheGl 319

1152 TTATCGACCCAGACTGAATAATGGAGAAAATATTTCCTTTATTGAC.... 1197
     :   :::   :::|||::::::|||:::|||::::::||| ' |||
319 nGlnLysGluHisLeuGlnThrGlyGlnAsnValThrPheAspAspSerT 336

1198 .....GAAGGAAAAGGCGAATTGATACTTACCAGCAACATCAATCAAGGT 1242
          :::|||||||||   |||||||||   ::::::||||||||||||
336 hrAsnAsnGlyLysGlyThrLeuIleLeuAspAspHisIleAsnGlnGly 352

1243 GCTGGAGGATTATATTTCCAAGGAGATTTTACGGTCTCGCCTGAAAATAA 1292
     |||||||||||||:::|||:::|||::::::   |||:::   ::::::::
353 AlaGlyGlyLeuPhePheLysGlyAsnTyrGluValLysGlyLysThrAs 369

1293 CGAA...ACTTGGCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCG 1339
     ::::   ||||||   |||:::|||:::   :::::::|||   :::   |
369 pAspIleThrTrpValGlyGlyGlyIleAspValAlaGluGlyLysGluV 386

1340 TTACTTGGAAAGTAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGC 1389
     ||:::|||||||||||::::::      :::|||:::|||:::|||||||||
386 alValTrpLysValHisAsnProGluLysAspHisLeuAlaLysIleGly 402

1390 AAAGGCACGCTGCACGTTCAAGCCAAAGGGGAAAACCAAGGCTCGATCAG 1439
     ||||||||||||   |||::::::|||||||:::|||:::|||||:::::
403 LysGlyThrLeuIleValGluGlyLysGlyAsnAsnLysGlySerLeuLy 419

1440 CGTGGGCGACGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCA 1489
     :|||||||||||||||||||:::|||   ||||||::::::   :::|||
419 sValGlyAspGlyThrValValLeuLysGlnGlnThrAsnGlySerGly. 435

1490 AAAAACAAGCCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTG 1539
       ::::::|||||:::::::::|||:::|||||||||||||:::||||||
436 ..GluHisAlaPheAlaSerValGlyIleValSerGlyArgSerThrVal 451

1540 CAACTGAATGCCGATAATCAGTTCAACCCCGACAAACTCTATTTCGGCTT 1589
       ||||||   |||:::|||   :::|||:::::::::||||||||||
452 ValLeuAsnAspAspLysGlnValAspProAsnSerIleTyrPheGlyPh 468

1590 TCGCGGCGGA 1599
     ||||||||||
468 eArgGlyGly 471
```

```
seq_name: sp_bacteria:O68900

seq_documentation_block:
ID   O68900      PRELIMINARY;      PRT;  1295 AA.
AC   O68900;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   PET PRECURSOR.
GN   PET.
OS   Escherichia coli.
OG   Plasmid pO42.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
```

```
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=042;
RX   MEDLINE=98298033; PubMed=9632580;
RA   Eslava C., Navarro-Garcia F., Czeczulin J.R., Henderson I.R.,
RA   Cravioto A., Nataro J.P.;
RT   "Pet, an autotransporter enterotoxin from enteroaggregative
RT   Escherichia coli.";
RL   Infect. Immun. 66:3155-3163(1998).
DR   EMBL; AF056581; AAC26634.1; -.
DR   MEROPS; S06.004; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
KW   Signal; Plasmid.
FT   SIGNAL       1     52       POTENTIAL.
SQ   SEQUENCE   1295 AA;  139768 MW;  9C6AEF7F345AD429 CRC64;


alignment_scores:
          Quality:  672.50              Length:   1507
            Ratio:  0.919                Gaps:     67
 Percent Similarity:  48.573   Percent Identity: 22.362

alignment_block:
 US-09-303-518D-649 x O68900   ..

 Align seg 1/1  to: O68900  from: 1  to: 1295

     109 CTTCCCCAAGCCTGGGCGGGACACACTTATTTCGGCATCAACTACCAATA 158
         :::   :::|||||||||                              :::
      57 IleSerLysAlaTrpAla.........................ArgAs 64

     159 CTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAG 208
          |||   |||:::|||:::|||||||||   |||      |||:::
      64 pTyrLeuAspLeuAlaGlnAsnLysGlyValPheGlnProGlySerThrH 81

     209 ATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACA 258
           :::::::::      |||   |||      :::      |||:::
      81 isValLysIleLysLeuLysAspGly......ThrAspPheSerPhePro 95

     259 AAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCGGC 308
               |||:::   |||||||||   ::::::   ||||||:::|||::
      96 AlaLeuProValProAspPheSerSerAlaThrAlaAsnGlyAlaAlaTh 112

     309 ATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATA 358
         :   :::|||      |||   |||:::|||||||||||||
     112 rSerIleGlyGlyAlaTyrAlaValThrValAlaHisAsn.......... 125

     359 ACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTT 408
                       |||:::::::::::      :::::::::
     126 ................AlaLysAsnLysSerSerAlaAsnTyrGln... 135

     409 ACTTATAAAATTGTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCA 458
         ||||||      ::::::   ::::::   :::      |||   |||::
     136 ThrTyrGlySerThrGlnTyrThrGlnIleAsnArgMetThrThrGlyAs 152

     459 TCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAATTTGTCACAG 508
         :            |||:::   :::   ||||||:::|||:::|||:::::
     152 n............AspPheSerIleGlnArgLeuAsnLysTyrValValG 165

     509 ATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGGAAATATATC 558
         ::::::::   ::::::   ||||||:::            :::|||
     165 luThrArgGlyAlaAsp...ThrSerPhe............AsnTyrAsn 176

     559 GATCAAAAT...AATTACCCTGACCGT..................... 582
         ::::::|||   |||      ||||||
     177 GluAsnAsnGlnAsnIleIleAspArgTyrGlyValAspValGlyAsnGl 193

     583 ..................GTTCGTATTGGGGCAGGCAGG.......... 603
                            |||:::|||:::|||:::
     193 yLysLysGluIleIleGlyPheArgValGlySerGlyAsnThrThrPheS 210

     604 ................CAATATTGGCGATCTGATGAAGATGAGCCCAAT 636

                         |||   :::::::::|||
     210 erGlyIleLysThrSerGlnThrTyrGlnAlaAsp............Leu 222

     637 AACCGCGAAAGTTCATATCATATTGCAAGTGCGTATTCTTGGCTCGTTGG 686
               |||  :::|||||||::::::      :::      |||||
     223 LeuSerAlaSerLeuPheHisIleThrAsnLeuArgAlaAsnThrValGl 239

     687 TGGCAATACCTTTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAG 736
         |||||||
     239 yGlyAsn.......................................... 241

     737 GTAGTGAAAAAATTAAACATAGCCCATATGGTTTTTTACCAACAGGAGGC 786
               |||:::::::::::::      :::::::::   :::
     242 ........LysValGluTyrGluAsnAspSerTyrPheThrAsnLeuThr 255

     787 TCATTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAAAAGCA 836
         :::   ||||||||||||||||   ::::::::::::|||   ::::::::
     256 ThrAsnGlyAspSerGlySerGlyValTyrValPheAspAsnLysGluAs 272

     837 AAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATATAGGAA 886
         :|||||||
     272 pLysTrp.......................................... 274

     887 AAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATGAAATC 936
                                                       :::
     275 .............................................Val 275

     937 TTTGCTGGAGATACCCATTCAGTATTCTACGAACCACGTCAAAATGGGAA 986
         :::   |||   ||||||::::::::::                |||||||||
     276 LeuLeuGlyThrThrHisGlyIleIle............GlyAsnGlyLy 288

     987 ATACTCTTTTAACGACGATAATAATGGCACAGGAAAAATCAATGCCAAAAC 1036
         |                                  |||   |||      :::
     288 s.........................ThrGlnLysThrTyrValThrP 296

    1037 ATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAACCGTTCAATTG 1086
               ::::::      |||:::|||||||          :::|||
     296 roPheAspSerLysThrThrAsnGluLeuLysGlnLeuPheIleGlnAsn 312

    1087 TTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTTTATCATGCTGC 1136
           |||::::::   :::   ::::::                    :::
     313 ValAsnIleAspAsnAsnThrAlaThrIleGlyGlyGlyLysIleThrIl 329

    1137 AGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGGAGAAAATATTT 1186
          |||::::::::::::          :::   :::|||::::::|||   :
     329 eGlyAsnThrThrGlnAspIleGluLysAsnLysAsnAsnGlnAsnLysA 346

    1187 CCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCAGCAACATCAAT 1236
         ::::::::   :::|||   |||::::::   |||   :::|||::::::
     346 spLeuValPheSerGlyGlyGlyLysIleSerLeuLysGluAsnLeuAsp 362

    1237 CAAGGTGCTGGAGGATTATATTTCCAAGGAGAT......TTTACGGTCTC 1280
           |||   ||||||:::   |||:::   :::      :::|||||||||
     363 LeuGlyTyrGlyGlyPheIlePheAspGluAsnLysLysTyrThrValSe 379

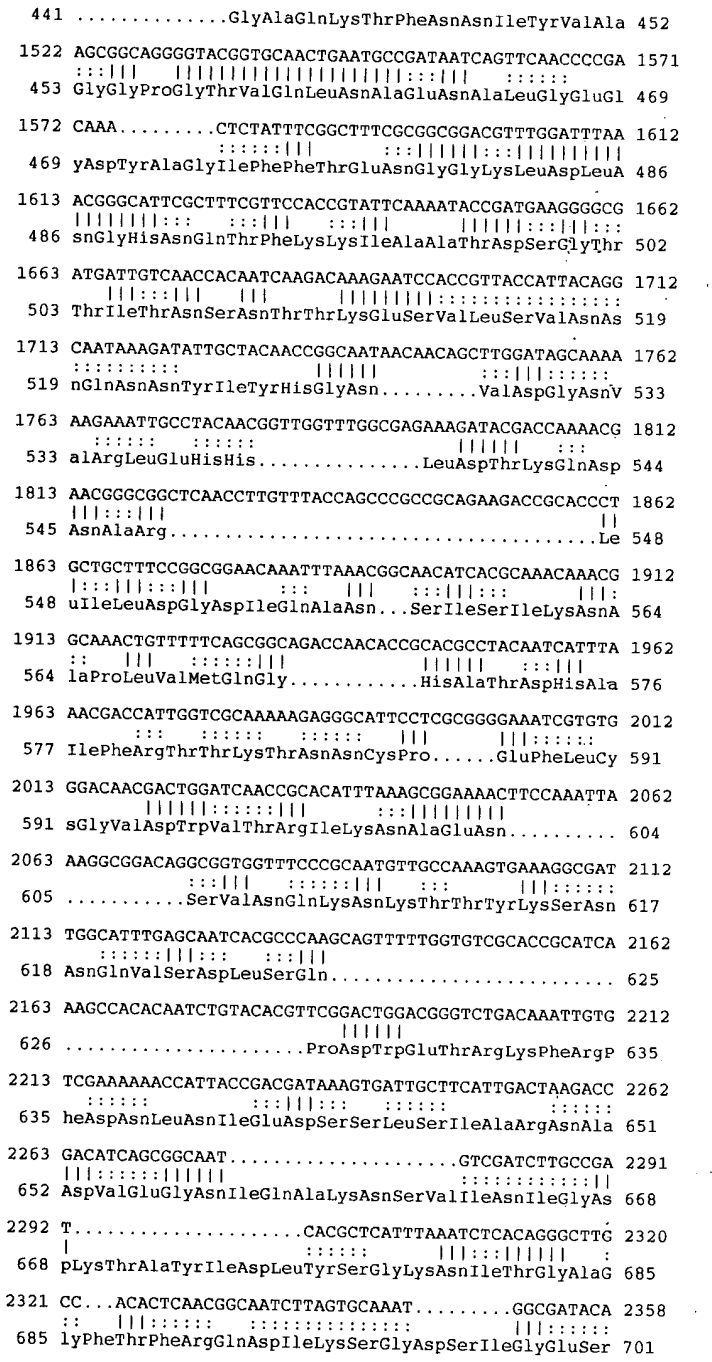
    1281 GCCTGAA......AATAACGAAACTTGGCAAGGCGCGGGCGTTCATATCA 1324
         |   |||      ||||||   |||::::::|||||||||:::   |||:
     379 rAlaGluGlyAsnAsnAsnValThrPheLysGlyAlaGlyIleAspIleG 396

    1325 GTGAAGACAGTACCGTTACTTGGAAAGTAAACGGCGTGGCAAACGACCGC 1374
         :::::   |||||||||   |||:::::::::   ::::::||||||
     396 lyLysGlySerThrValAspTrpAsnIleLysTyrAlaSerAsnAspAla 412

    1375 CTGTCCAAAATCGGCAAAGGCACGCTGCACGTT...CAAGCCAAAGGGGA 1421
         |||   |||||||||:::|||:::|||:::|||   ||||||::::::
     413 LeuHisLysIleGlyGluGlySerLeuAsnValIleGlnAlaGlnAsnTh 429

    1422 AAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTCATTTTGGATCAGC 1471
          |||         :::::::::|||:::||||||||||||||||
     429 rAsn.........LeuLysThrGlyAsnGlyThrValIleLeu....... 440

    1472 AGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGAAATCGGCTTGGTC 1521
                       |||   ::::::::::|||:::::|||   ::::::
```


441 .....GlyAlaGlnLysThrPheAsnAsnIleTyrValAla 452
1522 AGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATCAGTTCAACCCCGA 1571
453 GlyGlyProGlyThrValGlnLeuAsnAlaGluAsnAlaLeuGlyGluGl 469
1572 CAAA.....CTCTATTTCGGCTTTCGCGGCGGACGTTTGGATTTAA 1612
469 yAspTyrAlaGlyIlePhePheThrGluAsnGlyGlyLysLeuAspLeuA 486
1613 ACGGGCATTCGCTTTCGTTCCACCGTATTCAAAATACCGATGAAGGGGCG 1662
486 snGlyHisAsnGlnThrPheLysLysIleAlaAlaThrAspSerGlyThr 502
1663 ATGATTGTCAACCACAATCAAGACAAAGAATCCACCGTTACCATTACAGG 1712
503 ThrIleThrAsnSerAsnThrThrLysGluSerValLeuSerValAsnAs 519
1713 CAATAAAGATATTGCTACAACCGGCAATAACAACAGCTTGGATAGCAAAA 1762
519 nGlnAsnAsnTyrIleTyrHisGlyAsn.....ValAspGlyAsnV 533
1763 AAGAAATTGCCTACAACGGTTGGTTTGGCGAGAAAGATACGACCAAAACG 1812
533 alArgLeuGluHisHis.....LeuAspThrLysGlnAsp 544
1813 AACGGGCGGCTCAACCTTGTTTACCAGCCCGCCGCAGAAGACCGCACCCT 1862
545 AsnAlaArg.....Le 548
1863 GCTGCTTTCCGGCGGAACAAATTTAAACGGCAACATCACGCAAACAAACG 1912
548 uIleLeuAspGlyAspIleGlnAlaAsn...SerIleSerIleLysAsnA 564
1913 GCAAACTGTTTTTCAGCGGCAGACCAACACCGCACGCCTACAATCATTTA 1962
564 laProLeuValMetGlnGly.....HisAlaThrAspHisAla 576
1963 AACGACCATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGAAATCGTGTG 2012
577 IlePheArgThrThrLysThrAsnAsnCysPro.....GluPheLeuCy 591
2013 GGACAACGACTGGATCAACCGCACATTTAAAGCGGAAAACTTCCAAATTA 2062
591 sGlyValAspTrpValThrArgIleLysAsnAlaGluAsn..... 604
2063 AAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAAGTGAAAGGCGAT 2112
605 .....SerValAsnGlnLysAsnLysThrThrTyrLysSerAsn 617
2113 TGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGTGTCGCACCGCATCA 2162
618 AsnGlnValSerAspLeuSerGln..... 625
2163 AAGCCACACAATCTGTACACGTTCGGACTGGACGGGTCTGACAAATTGTG 2212
626 .....ProAspTrpGluThrArgLysPheArgP 635
2213 TCGAAAAAACCATTACCGACGATAAAGTGATTGCTTCATTGACTAAGACC 2262
635 heAspAsnLeuAsnIleGluAspSerSerLeuSerIleAlaArgAsnAla 651
2263 GACATCAGCGGCAAT.....GTCGATCTTGCCGA 2291
652 AspValGluGlyAsnIleGlnAlaLysAsnSerValIleAsnIleGlyAs 668
2292 T.....CACGCTCATTTAAATCTCACAGGGCTTG 2320
668 pLysThrAlaTyrIleAspLeuTyrSerGlyLysAsnIleThrGlyAlaG 685
2321 CC...ACACTCAACGGCAATCTTAGTGCAAAT.....GGCGATACA 2358
685 lyPheThrPheArgGlnAspIleLysSerGlyAspSerIleGlyGluSer 701
2359 CGTTATACAGTCAGCCACAACGCCACCCAAAACGGCAACCTTAGCCTCGT 2408
702 LysPheThrGlyGlyIleMetAlaThr...AspGlySerIleSerIleGl 717
2409 GGGCAATGCCCAAGCAACATTTAATCAAGCC..... 2439
717 yAspLysAlaIleValThrLeuAsnThrValSerSerLeuAspArgThrA 734
2440 .....ACATTAAACGGCAACACATCGGCTTCGGGCAATGCTTCATTTAAT 2484
734 laLeuThrIleHisLysGlyAlaAsnValThrAlaSerSerSerLeuPhe 750
2485 CTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGGCAACGC 2534
751 ThrThrSerAsnIleLysSerGlyGlyAspLeuThrLeuThrGly..... 765
2535 TAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCCTAGCCG 2584
765 765
2585 ATAAGGCAGTATTCCATTTTGAAAGCAGCCGCTTTACCGGACAAATCAGC 2634
766 .....AlaThrGluSerThrGlyGluIleThr 774
2635 GGCGGCAAGGATACGGCATTACACTTAAAAGACAGCGAATGGACGCTGCC 2684
775 .....ProSerMetPheTyrAlaAl 781
2685 GTCAGGCACGGAATTA.....GGCAATTTAAACCTTGACAACG 2722
781 aGlyGlyTyrGluLeuThrGluAspGlyAlaAsnPheThrAlaLysAsnG 798
2723 CCACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGCGCAA 2772
798 lnAlaSerValThrGlyAspIleLysSerGluLysAlaAlaLysLeuSer 814
2773 ACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGCGCCG 2822
815 PheGlySerAla...AspLysAspAsnSerAlaThrArgTyrSerGlnPh 830
2823 TTCCCTATTATCCGTTACACCGCCAACTTCGGTAGAATCCCGTTTCAACA 2872
830 eAlaLeu..... 832
2873 CGCTGACGGTAAACGGCAAATTGAACGGTCAGGGAACATTCCGCTTTATG 2922
833 .....AlaMetLeuAspGlyPheAspThr..... 840
2923 TCGGAACTCTTCGGCTACCGCAGCGACAAATTGAAGCTGGCGGAAAGTTC 2972
841 .....SerTyrGlnGly...SerIleLysAlaAlaGlnSerSe 852
2973 CGAAGGCACTTACACCTTGGCGGTCAACAAT.....ACCG 3007
852 r.....LeuAlaMetAsnAsnAlaLeuTrpLysValThrG 864
3008 GCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGAC 3057
864 lyAsn.....SerGlu 867
3058 AACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGT 3107
868 LeuLysLysLeuAsnSerThrGlySerMetValLeuPheAsn..... 881
3108 CGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCC 3157
882 .....GlyGlyLysA 885
3158 GCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAG 3207
885 snIlePheAsnThrLeuThrValAspGluLeuThr..... 896

```
3208 GCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGC 3257
     ::::::   :::      :::   :::   |||:::|||:::   |||
 897 ThrSerAsnSerAlaPheValMetArgThrAsnThrGlnGlnAlaAspGl 913

3258 GCTGATTGCG...........GCCGGGCGCGATGCCGTCGAAAAGACAG 3295
      ||||||:::              |||   :::   :::      ::::
 913 nLeuIleValLysAsnLysLeuGluGlyAlaAsnAsnLeuLeuLeuValA 930

3296 AAAGCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATT 3345
     ::   :::   :::   :::::::::   :::|||   |||:::   :::
 930 spPheIleGluLysLysGlyAsnAspLysAsnGlyLeuAsnIleAspLeu 946

3346 ATGCAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGC 3395
     ::::::|||   |||              :::::::::||||||:::
 947 ValLysAlaProGlu.............AsnThrSerLysAspValPh 958

3396 CTTGGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCC 3445
               :::   |||:::::::|||       :::   :::   |
 958 e.........LysThrGluThrGlnThrIleGlyPheSerAspValThrP 972

3446 CCCGCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCG 3495
     ||:::   ::::::   ::::::|||
 972 roGluIleLysGlnGlnGluLysAsp..................... 980

3496 CAGCCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTT 3545
                                                   |||
 981 .............................................GlyLy 982

3546 GAGTGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAAT 3595
      |||   ::::::   |||      ::::::   |||      |||
 982 sSerValTrpThrLeuThrGlyTyrLysThrValAlaAsnAlaAsp.... 997

3596 TAGACCGCGTATTTGCCGAAGACCGCCGCAACGCCGTTTGGACAAGCGGC 3645
                |||      :::   ::::::::::   ::::::|||
 998 ..............AlaAlaLysLysAlaThrSerLeuMetSerGlyGly 1009

3646 ATCCGGGACACCAAACACTACCGTTCGCAAGATTTCCGCGCCTACCGCCA 3695
               |||   :::   ::::::      :::      :::::
1010 Tyr.........LysAlaPheLeuAlaGluValAsnAsnLeuAsnLysAr 1023

3696 ACAAACCGACCTGCGCCAAATCGGTATGCAGAAAAACCTC.......... 3735
     ::::   ||||||||::::|||:::   :::   :::
1023 gMetGlyAspLeuArgAspIleAsnGlyGluAlaGlyAlaTrpAlaArgI 1040

3736 ........GGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACC 3777
             |||:::|||   :::|||      ||||||   |||   |||
1040 leMetSerGlyThrGlySerAlaGlyGlyGlyPheSerAspAsnTyrThr 1056

3778 ..................GAAAACACCTTCGAC...GACGGCATCGGCAA 3806
                        :::|||      :::   ||||||:::
1057 HisValGlnValGlyAlaAspAsnLysHisGluLeuAspGlyLeuAspLe 1073

3807 CTCGGCACGGCTTGCCCACGGCGCCGTTTTCGGGCAATACGGCATCGACA 3856
        :::   ::::::         :::   ::::::   |||   |||
1073 uPheThrGlyValThrMetThrTyrThrAspSerHisAlaGlySerAspA 1090

3857 GGTTCTAC....................ATCGGCATCAGCGCGGGCGCG 3885
       |||                              |||:::   |||:::|||
1090 laPheSerGlyGluThrLysSerValGlyAlaGlyLeuTyrAlaSerAla 1106

3886 GGTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGGCAAAATCCGCCG 3935
       |||:::||||||:::       |||   ||||||
1107 MetPheGluSerGlyAlaTyrIleAspLeuIleGly.............. 1118

3936 CCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTTCGGCG 3985
      :::   :::|||:::       :::   :::|||   |||   |||:::|
1119 .LysTyrValHisHis......AspAsnGluTyrThrAlaThrPheAlaG 1133

3986 GATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCCAAAAA 4035
     ||:::|||                        ||||||
1133 lyLeuGly....................ThrArg.............. 1137

4036 GCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTTGCATT 4085
```

```
                                                   :::
        ||||||                                   SerSe 1141
1138 ...AspTyr.....................................

4086 CAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACCGGCGC 4135
      :::   :::   ||||||   ::::::   |||   :::       ::::
1141 rHisSerTrpTyrAlaGlyAlaGluValGlyTyrArgTyrHisValThrA 1158

4136 AACACATTTCCATCACGCCTTATTTGAGCCTGTCCTATACCGATGCCGCT 4185
     ::         |||   |||      :::|||   |||      |||:::
1158 spSerAlaTrpIleGluProGlnAlaGluLeuValTyr...GlyAlaVal 1173

4186 TCGGGCAAAGTCCGAACA.............CGCGTCAATACCGCCGTATT 4223
     |||||||||      :::              :::|||   ::::::
1174 SerGlyLysGlnPheSerTrpLysAspGlnGlyMetAsnLeuThrMetLy 1190

4224 GGCTCAGGATTTC................................GGCA 4240
        :::|||||||                                 ||||
1190 sAspLysAspPheAsnProLeuIleGlyArgThrGlyValAspValGlyL 1207

4241 AAACCCGCAGTGCG...GAATGGGGCGTAAACGCCGAAATC...AAAGGT 4284
     ||:::   |||:::   :::|||   |||:::|||:::         |||
1207 ysSerPheSerGlyLysAspTrpLysValThrAlaArgAlaGlyLeuGly 1223

4285 TTCACGCTGTCCCTCCACGCT 4305
     :::   ::::::|||   |||
1224 TyrGlnPheAspLeuPheAla 1230
```

seq_name: sp_bacteria:O69745

seq_documentation_block:

```
ID   O69745      PRELIMINARY;      PRT;   190 AA.
AC   O69745;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
OS   Neisseria lactamica.
OC   Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX   NCBI_TaxID=486;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NL3293;
RA   Sayers J.R., Vitowski S., Read R.;
RT   "Strains of Neisseria lactamica carry an IgA1 protease gene homologue
RT   and some express IgA1 protease activity.";
RL   Submitted (OCT-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AJ001738; CAA04963.1; -.
DR   MEROPS; S06.001; -.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     190    190
SQ   SEQUENCE   190 AA;  21016 MW;  4A524869DB83BAF1 CRC64;
```

alignment_scores:

```
            Quality:  660.50              Length:    192
              Ratio:  4.027                 Gaps:      4
 Percent Similarity:  85.417   Percent Identity: 66.667
```

alignment_block:

```
 US-09-303-518D-649 x O69745   ..

 Align seg 1/1  to: O69745  from: 1  to: 190

     205 AAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAAT 254
         ||||||||||||||||||||||||||||||||||||||||||||||||||
       1 LysAspIleGluValTyrAsnLysLysGlyGluLeuValGlyLysSerMe 17

     255 GACAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGG 304
         ||||||||||||||||||||||||||||||||||||||||||||||||||
      17 tThrLysAlaProMetIleAspPheSerValValSerArgAsnGlyValA 34

     305 CGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGC 354
```

```
      ||||||||||||||||||||||||||||||||||||||||||||||||||||   |||
       34 laAlaLeuValGlyAspGlnTyrIleValSerValAlaHisAsnValGly 50

      355 TATAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCG 404
          |||::::::|||||||||||||||||||||||   ||||||||||||||||
       51 TyrArgAspValAspPheGlyAlaGluGlySerAsnProAspGlnHisAr 67

      405 TTTTACTTATAAAATTGTGAAACGGAATAATTATAAAGCAGGGACTAAAG 454
          ||||:::|||||||||:::|||||||||||||||||||   :::   :::
       67 gPheSerTyrLysIleAlaLysArgAsnAsnTyrLys...AsnAspGluT 83

      455 GCCATCCTTATGGCGGCGATTATCATATGCCGCGTTTGCATAAATTTGTC 504
            |||||||||      |||||||||   ||||||||||||||||||||
       83 hrHisProTyrGluLysAspTyrHisAsnProArgLeuHisLysPheVal 99

      505 ACAGATGCAGAACCTGTTGAAATGACCAGTTATATGGATGGGCGGAAATA 554
          |||:::|||   |||:::::|||||||||   ||||||||||:::|||||
      100 ThrGluAlaAlaProIleAspMetThrSerAspMetAspGlyAsnLysTy 116

      555 TATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGCAGGC 604
          |   |||:::::::::::||||||:::|||||||||||||||||:::|||   |
      116 rThrAspArgThrLysTyrProGluArgValArgIleGlySerGlyTrpG 133

      605 AATATTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATAT 654
          ||:::|||||||:::|||:::|||:::   :::
      133 lnPheTrpArgAsnAspGlnAspLysGlyAsp..................... 143

      655 CATATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCTTTGCACA 704
          ::::::|||:::||||||   :::|||::::::|||||||||      ||
      144 GlnValAlaGlyAlaTyrHisTyrLeuThrAlaGlyAsnThrHisAsnGl 160

      705 AAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAAC 754
          |:::|||:::|||||||||      :::|||:::::::::   ::::::|
      160 nGlyGlyAlaGlyGlyGlyTrpSerSerLeuSerGlyAsp...ValArgH 176

      755 ATAGC...CCATATGGTTTTTTACCA 777
          ||:::      ||||||   :::|||
      176 isAlaGlyAsnTyrGlyProIlePro 184

seq_name: sp_bacteria:Q48206

seq_documentation_block:
ID   Q48206      PRELIMINARY;      PRT;   492 AA.
AC   Q48206;
DT   01-NOV-1996 (TrEMBLrel. 01, Created)
DT   01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   IGA1 PROTEASE (FRAGMENT).
GN   IGA.
OS   Haemophilus influenzae.
OC   Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC   Haemophilus.
OX   NCBI_TaxID=727;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=DA66;
RX   MEDLINE=95302961; PubMed=7783620;
RA   Lomholt H., Poulsen K., Mogens K.;
RT   "Comparative characterization of the iga gene encoding IgA1 protease
RT   in Neisseria meningitidis, Neisseria gonorrhoeae and Haemophilis
RT   influenzae.";
RL   Mol. Microbiol. 15:495-506(1995).
DR   EMBL; X82467; CAA57850.1; -.
DR   MEROPS; S06.001; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
KW   Protease.
FT   NON_TER       1      1
FT   NON_TER     492    492
SQ   SEQUENCE   492 AA;  54751 MW;  6FF93038E7AD5984 CRC64;

alignment_scores:
            Quality:  657.50              Length:    534
              Ratio:  2.156                 Gaps:     18
 Percent Similarity:  57.116   Percent Identity: 32.397

alignment_block:
 US-09-303-518D-649 x Q48206   ..

 Align seg 1/1  to: Q48206  from: 1  to: 492

      202 GCGAAAGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATC 251
          |||   ::::::|||||||   :::||||||::::::   :::|||:::
        1 AlaThrAsnValGluValArgAspLysLysAsnAsnAsnLeuGlySerVa 17

      252 AATGACAAAA...GCCCCGATGATTGATTTTTCTGTGGTGTCG...CGTA 295
            :::   |||      |||||||||||||||||||||||:::      :
       17 lLeuProLysAspIleProMetIleAspPheSerValValAspValAspL 34

      296 ACGGCGTGGCGGCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACAT 345
          ::   :::|||:::|||:::::::   ||||||:::|||:::|||   |||
       34 ysArgIleAlaThrLeuIleAsnProGlnTyrValValGlyValLysHis 50

      346 AACGGC...GGCTATAACAACGTTGATTTTGGT..................... 375
             |||   |||   :::::::::   ||||||
       51 ValGlyAsnGlyValGlyGluLeuHisPheGlyAsnLeuAsnGlyAsnTr 67

      376 ....GCGGAAGGAAGAAATCCCGATCAACATCGTTTTACT.......... 411
                ::::      |||      |||||||||   :::
       67 pAsnProLysPheGlyAsnSerIleGlnHisArgAspValSerTrpGluG 84

      412 ........TATAAAATTGTGAAACGGAATAATTATAAAGCAGGGACTAAA 453
                  |||      |||::::::||||||:::::::::      :::
       84 luAsnArgTyrTyrThrValGluLysAsnAsnPheSerSerGluLeuAsn 100

      454 GGCCATCCT..................................... 462
          |||
      101 GlyLysThrGlnAsnAsnGluLysAspLysGlnTyrThrSerAsnLysLy 117

      463 ..................TATGGC..................... 468
                           ||||||
      117 sAspValProSerGluLeuTyrGlyGlnAlaLeuValLysGluGlnGlnA 134

      469 ..............GGCGATTATCATATGCCGCGTTTGCATAAATTTGTC 504
                       ||||||:::|||||||||||   |||||||||
      134 snGlnLysArgArgGluAspTyrTyrMetProArgLeuAspLysPheVal 150

      505 ACAGATGCAGAACCTGTTGAA...ATGACCAGTTATATGGATGGGCGGAA 551
          |||::::::   |||:::|||      |||:::      |||:::
      151 ThrGluValAlaProIleGluAlaSerThrThrSerSerAspAlaGlyTh 167

      552 ATATATCGATCAAAATAATTACCCTGACCGTGTTCGTATTGGGGCAGGCA 601
            |||   |||||||||:::||||||      ||||||:::|||:::|||
      167 rTyrAsnAspGlnAsnLysTyrProAlaPheValArgLeuGlySerGlyS 184

      602 GGCAATAT.....................TGGCGATCTGATGAAGATGAG 630
            |||:::                     ::::::         :::|||
      184 erGlnPheIleTyrLysLysGlySerHisTyrGluLeuIleLeuGluGlu 200

      631 CCCAATAACCGCGAAAGTTCATATCAT..................... 657
            |||::::::::::::      |||
      201 LysAsnGluLysLysGluIleIleHisArgTrpAspValGlyGlyAspAs 217

      658 ..........ATTGCAAGTGCGTATTCTTGGCTCGTTGGTGGCAATACCT 697
                   :::::::::||||||::::::   ::::::|||:::   :
      217 nLeuLysLeuValGlyAsnAlaTyrThrTyrGlyIleAlaGlyThrProT 234

      698 TTGCACAAAATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAA 747
          ::      |||   :::      |||   :::::::::|||::::::
      234 yrLysValAsnHisThrAspAspGlyLeuIleGlyPheGlyAspSerThr 250

      748 ATTAAACATAGCCCA................................ 762
                ||:::
      251 GluAspHisAsnAspProLysGluIleLeuSerArgLysProLeuThrAs 267
```

```
 763 .TATGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAA 811
     |||:::   |||                  ||||||||||||||||||:
 267 nTyrAlaValLeu.................GlyAspSerGlySerProL 278

 812 TGTTTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTG 861
     ::|||:::||||||   :::|||:::|||||||||::   |||:::
 278 euPheValTyrAspLysSerLysGluLysTrpLeuPheLeuGlyAlaTyr 294

 862 CAAACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCG 911
     :::      :::   |||      ||||||                :::
 295 AspPheTrpGlyGlyTyrLysLysLysSerTrpGlnGluTrpAsnIleTy 311

 912 TAAAGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTAT 961
      |||      |||   ::::::|||:::      |||:::   :::::::
 311 rLysProGlnPheAlaGluAsnIleLeuLysLysAspSerAlaGlyLeuL 328

 962 TCTACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAAT 1011
     ::         :::   |||   :::|||:::::::::::::   :::|||
 328 eu.........LysGlyAsnThrGlnTyrAsnTrpThrSerLysGlyAsn 341

1012 GGCACAGGAAAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAG 1061
                                                    :::
 342 ..............................................ThrSe 343

1062 ATTAAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAG 1111
      |||   :::   |||   :::   ::::::|||:::|||   :::
 343 rLeuIleSerGlyThrSerGluSerLeuSerValAspLeuValAsp.... 358

1112 CAAGAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCC 1161
                                                   :::
 359 ............................................AsnLys 360

1162 AGACTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGA 1211
     :::||||||:::|||:::|||::::::|||      ||||||:::|||::
 361 AsnLeuAsnHisGlyLysAsnValThrPhe......GluGlySerGlyAs 375

1212 ATTGATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCC 1261
     :|||   |||::::::||||||:::|||||||||||||||||||:::|||:
 375 nLeuThrLeuAsnAsnAsnIleAspGlnGlyAlaGlyGlyLeuPhePheG 392

1262 AAGGAGATTTTACGGTC...TCGCCTGAAAATAACGAAACTTGGCAAGGC 1308
     ::|||||:::   |||   :::   ::::::|||   ||||||:::|||
 392 luGlyAspTyrGluValLysGlyThrSerGluAsnThrThrTrpLysGly 408

1309 GCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGG 1358
     ||||||:::   ::::::|||   :::|||||||   |||||||||:::::
 409 AlaGlyIleSerValAlaGluGlyLysThrValLysTrpLysValHisAs 425

1359 CGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTC 1408
     :         |||||||||:::|||||||||||||||   |||   |||:
 425 nProGlnPheAspArgLeuAlaLysIleGlyLysGlyLysLeuIleValG 442

1409 AAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTC 1458
     ::::::::|||:::|||:::||||||:::::|||||||||||||||||||
 442 luGlyArgGlyAspAsnLysGlySerLeuLysValGlyAspGlyThrVal 458

1459 ATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGA 1508
     :::|||   ||||||:::                ::::::||||||:::::
 459 ValLeuLysGlnGlnThrThrThr......GlyGlnHisAlaPheAlaSe 473

1509 AATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAATGCCGATAATC 1558
     ::::|||:::||||||||||||:::||||||   ||||||   |||||||
 473 rValGlyIleValSerGlyArgSerThrValValLeuAsnAspAspAsnG 490

1559 AG 1560
     ||
 490 ln 490
```

seq_name: sp_bacteria:Q9EZE7

seq_documentation_block:
ID Q9EZE7 PRELIMINARY; PRT; 1305 AA.

AC Q9EZE7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE ENTEROTOXIN ESPC.
GN ESPC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=E2348/69;
RX MEDLINE=20569196; PubMed=11119520;
RA Mellies J.L., Navarro-Garcia F., Okeke I., Frederickson J.,
RA Nataro J.P., Kaper J.B.;
RT "espC pathogenicity island of enteropathogenic Escherichia coli
RT encodes an enterotoxin.";
RL Infect. Immun. 69:315-324(2001).
DR EMBL; AF297061; AAG37043.1; -.
DR MEROPS; S06.UPW; -.
DR InterPro; IPR000710; IgA_S6.
DR Pfam; PF02395; IGA1; 1.
DR PRINTS; PR00921; IGASERPTASE.
SQ SEQUENCE 1305 AA; 140782 MW; 8ECE9ED225C31F5C CRC64;

alignment_scores:
Quality: 641.50 Length: 1478
Ratio: 0.882 Gaps: 68
Percent Similarity: 49.188 Percent Identity: 23.139

alignment_block:
US-09-303-518D-649 x Q9EZE7 ..

Align seg 1/1 to: Q9EZE7 from: 1 to: 1305

```
 160 TATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAAAGA 209
     |||   |||:::|||:::|||||||||   |||   :::||||||   ::
  66 TyrLeuAspLeuAlaGlnAsnLysGlyValPheLysAlaGlyAlaThrAs 82

 210 TATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGACAA 259
     ::::::::::      |||:::|||:::            ::::::   :
  82 nValSerIleGlnLeuLysAsnGlyGlnThrPhe......AsnPheProA 97

 260 AAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCGGCA 309
     :::::|||:::   |||||||||   :::|||:::::::|||:::::::::
  97 snValProIleProAspPheSerProAlaSerAsnLysGlyAlaThrThr 113

 310 TTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTATAA 359
        :::|||      |||   |||:::::::::|||||||||
 114 SerIleGlyGlyAlaTyrSerValThrAlaThrHisAsnGly........ 127

 360 CAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGATCAACATCGTTTTA 409
                                                      |
 128 ..............................................ThrT 129

 410 CTTATAAAATTGTGAAACGGAATAATTATAAAGCAGGGACTAAAGGCCAT 459
     ||:::      ::::::   :::|||:::      |||   :::::::::
 129 hrHisHisAlaIleSerThrGlnAsnTrp......GlyGlnSerSerTyr 143

 460 CCTTAT..............GGCGGCGATTATCATATGCCGCGTTTGCA 494
         |||              :::|||||||:::   :::   ||||||
 144 LysTyrIleAspArgMetThrAsnGlyAspPheAlaValThrArgLeuAs 160

 495 TAAATTTGTCACAGATGCAGAACCTGTTGAA.................A 526
      |||||||||:::::::::      |||:::                 :
 160 pLysPheValValGluThrThrGlyValLysAsnSerValAspPheSerL 177

 527 TGACCAGTTATATGGATGGGCGGAAATATATCGATCAAAATAATTACCCT 576
     :::::|||:::         :::::::|||      :::   |||
 177 euAsnSerHisAspAlaLeuGluArgTyrGlyValGluIleAsnGlyGlu 193
```

```
 577 GACCGTGTT.........CGTATTGGGGCAGGCAGGCAATATTGGCGATC 617
 194 LysLysIleIleGlyPheArgValGlyAlaGlyThrThrTyr........ 207

 618 TGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATATTGCAAGTG 667
 208 .............ThrValGlnAsnGlyAsnThrTyrSerThrGlyGlnV 220

 668 CGTATTCT...TGGCTCGTTGGTGGCAATACCTTTGCACAAAATGGATCA 714
 220 alTyrAsnProLeuLeuSerAlaSerMetPheGlnLeuAsnTrpAsp 236

 715 GGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACATAGCCCATA 764
 237 AsnLysArgProTyrAsn.................AsnThrThrProPh 247

 765 TGGTTTTTTACCAACAGGAGGCTCATTTGGCGACAGTGGCTCACCAATGT 814
 247 eTyrAsnGluThrThrGly.........GlyAspSerGlySerGlyPheT 261

 815 TTATCTATGATGCCCAAAAGCAAAAGTGGTTAATTAATGGGGTATTGCAA 864
 261 yrLeuTyrAspAsnValLysLysGluTrpValMetLeuGlyThrLeu... 276

 865 ACGGGCAACCCCTATATAGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAA 914
 276 ..................................................... 276

 915 AGATTGGTTCTATGATGAAATCTTTGCTGGAGATACCCATTCAGTATTCT 964
 277 .......PheGlyIleAlaSerSerGlyAlaAspValTrpSerIleLeuA 291

 965 ACGAACCACGTCAAAATGGGAAATACTCTTTTAACGACGATAATAATGGC 1014
 291 snGlnTyrAspGluAsn.............................. 296

1015 ACAGGAAAAATCAATGCCAAACATGAACACAATTCTCTGCCTAATAGATT 1064
 297 .......................ThrValAsnGlyLeuLysAsnLysPh 305

1065 AAAAACACGAACCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAA 1114
 305 e...ThrGlnLysValGlnLeuAsnAsnAsnThrMetSer......LeuA 319

1115 GAGAACCTGTTTATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGA 1164
 319 snSerAspSerPheThrLeuAlaGlyAsnAsnThrAlaValGluLysAsn 335

1165 CTGAATAATGGAGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATT 1214
 336 AsnAsnAsnTyrLysAspLeuSerPhe......SerGlyGlyGlySerIl 350

1215 GATACTTACCAGCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAG 1264
 350 eAsnPheAspAsnAspValAsnIleGlySerGlyGlyLeuIlePheAspA 367

1265 GA......GATTTTACGGTCTCGCCTGAAAATAACGAAACTTGGCAAGGC 1308
 367 laGlyHisHisTyrThrValThr...GlyAsnAsnLysThrPheLysGly 382

1309 GCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAGTAAACGG 1358
 383 AlaGlyLeuAspIleGlyAspAsnThrThrValAspTrpAsnValLysGl 399

1359 CGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTGCACGTTC 1408
 399 yValValGlyAspAsnLeuHisLysIleGlyAlaGlyThrLeuAsnValA 416

1409 AAGCCAAAGGGGAAAACCAAGGCTCGATCAGCGTGGGCGACGGTACAGTC 1458
 416 snValSerGlnGlyAsn......AsnLeuLysThrGlyAspGlyLeuVal 430

1459 ATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAGCCTTTAGTGA 1508
```

```
 431 ValLeuAsn.....................SerAlaAsnAlaPheAspAs 440

1509 AATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAAT.........G 1549
 440 nIleTyrMetAlaSerGlyHisGlyValValLysIleAsnHisSerAlaA 457

1550 CCGATAATCAGTTCAACCCCGACAAA...CTCTATTTCGGCTTTCGCGGC 1596
 457 laLeuAsnGlnAsnAsnAspTyrArgGlyIlePhePheThrGluAsnGly 473

1597 GGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTATTCAAAA 1646
 474 GlyThrLeuAspLeuAsnGlyTyrAspGlnSerPheAsnLysIleAlaAl 490

1647 TACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAAGAATCCA 1696
 490 aThrAspIleGlyAlaLeuIleThrAsnSerAlaValGlnLys...AlaV 506

1697 CCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAATAACAAC 1746
 506 alLeuSerValAsnAsnGlnSerAsnTyrMetTyrHisGly......... 519

1747 AGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTGGCGAGAA 1796
 520 SerValSerGlyAsnThrGluIleAsnHis.............GlnPh 531

1797 AGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAGCCCGCCG 1846
 531 eAspThrGlnLysAsnAsnSerArg..................... 539

1847 CAGAAGACCGCACCCTGCTGCTTTCCGGCGGAACAAATTTAAACGGCAAC 1896
 540 .............LeuIleLeuAspGlyAsnValAspIleThrAsnAsp 551

1897 ATCACGCAAACAAACGGCAAACTGTTTTTCAGCGGCAGACCAACACCGCA 1946
 552 IleAsnIleLysAsnSerGlnLeuThrMetGlnGlyHisAlaThrSerHi 568

1947 CGCCTAC.......................................... 1953
 568 sAlaValPheArgGluGlyGlyValThrCysMetLeuProGlyValIleC 585

1953 ..................................................... 1953
 585 ysGluLysAspTyrValSerGlyIleGlnGlnGlnGluAsnSerAlaAsn 601

1954 AATCATTTAAACGACCATTGGTCGCAAAAAGAGGGCATTCCTCGCGGGGA 2003
 602 LysAsnAsnAsnThrAspTyrLysThrAsnAsnGlnValSerSerPheGl 618

2004 AATCGTGTGGGACAACGACTGGATCAACCGCACATTTAAAGCGGAAAACT 2053
 618 u.........GlnProAspTrpGluAsnArgLeuPheLysPheLysThrL 632

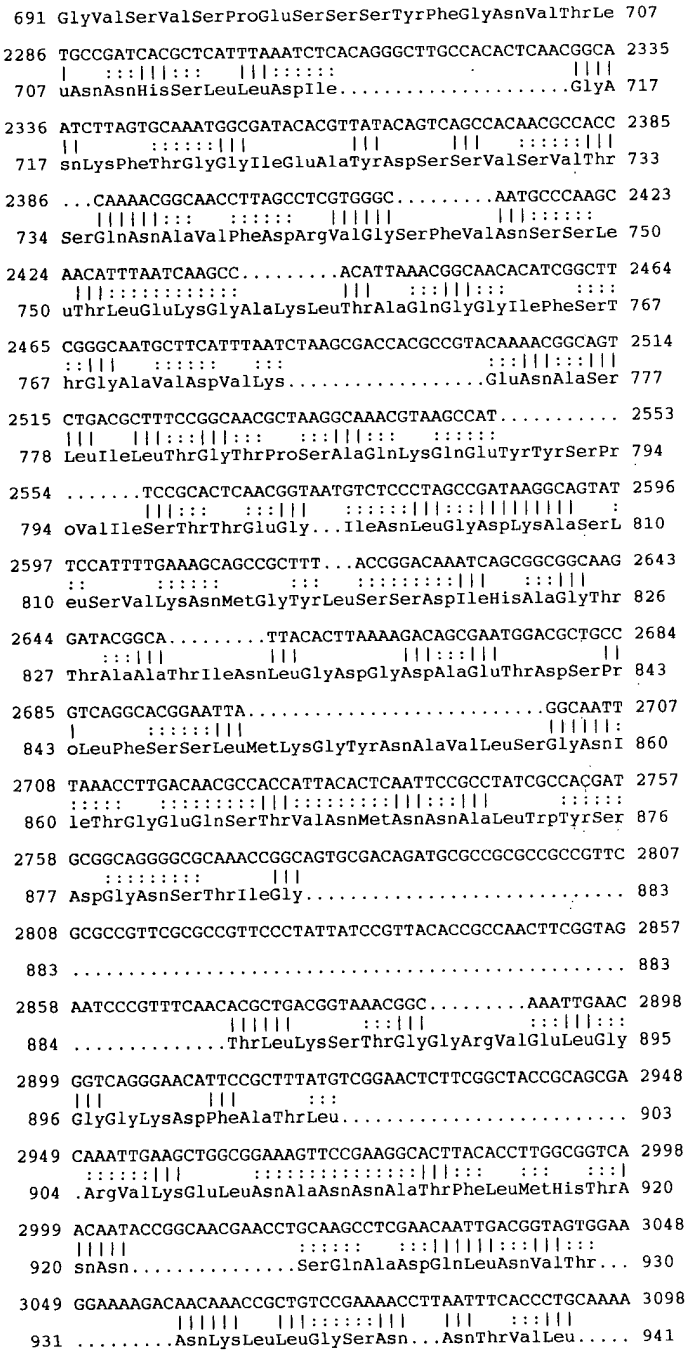
2054 TCCAAATTAAAGGCGGACAGGCGGTGGTTTCCCGCAATGTTGCCAAAGTG 2103
 632 euAsnLeuIleAsnSerAspPheIleValGlyArgAsn...AlaIleVal 647

2104 AAAGGCGATTGGCATTTGAGCAATCACGCCCAAGCAGTTTTTGGT..... 2148
 648 ValGlyAspIleSerAlaAsnAsnSerThrLeuSerLeuSerGlyLysAs 664

2149 .............GTCGCACCGCATCAAAGCCACACAATCTGTACACGTT 2185
 664 pThrLysValHisIleAspMetTyrAspGlyLysAsnIle......ThrG 679

2186 CGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACCGACGAT 2235
 679 lyAspGlyPheGlyPheArgGlnAspIleLys..............Asp 690

2236 AAAGTGATTGCTTCATTGACTAAGACCGACATCAGCGGCAATGTCGATCT 2285
```


691 GlyValSerValSerProGluSerSerSerTyrPheGlyAsnValThrLe 707
2286 TGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTCAACGGCA 2335
707 uAsnAsnHisSerLeuLeuAspIle.....GlyA 717
2336 ATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAACGCCACC 2385
717 snLysPheThrGlyGlyIleGluAlaTyrAspSerSerValSerValThr 733
2386 ...CAAAACGGCAACCTTAGCCTCGTGGGC.....AATGCCCAAGC 2423
734 SerGlnAsnAlaValPheAspArgValGlySerPheValAsnSerSerLe 750
2424 AACATTTAATCAAGCC.....ACATTAAACGGCAACACATCGGCTT 2464
750 uThrLeuGluLysGlyAlaLysLeuThrAlaGlnGlyGlyIlePheSerT 767
2465 CGGGCAATGCTTCATTTAATCTAAGCGACCACGCCGTACAAAACGGCAGT 2514
767 hrGlyAlaValAspValLys.....GluAsnAlaSer 777
2515 CTGACGCTTTCCGGCAACGCTAAGGCAAACGTAAGCCAT..... 2553
778 LeuIleLeuThrGlyThrProSerAlaGlnLysGlnGluTyrTyrSerPr 794
2554 .....TCCGCACTCAACGGTAATGTCTCCCTAGCCGATAAGGCAGTAT 2596
794 oValIleSerThrThrGluGly...IleAsnLeuGlyAspLysAlaSerL 810
2597 TCCATTTTGAAAGCAGCCGCTTT...ACCGGACAAATCAGCGGCGGCAAG 2643
810 euSerValLysAsnMetGlyTyrLeuSerSerAspIleHisAlaGlyThr 826
2644 GATACGGCA.....TTACACTTAAAAGACAGCGAATGGACGCTGCC 2684
827 ThrAlaAlaThrIleAsnLeuGlyAspGlyAspAlaGluThrAspSerPr 843
2685 GTCAGGCACGGAATTA.....GGCAATT 2707
843 oLeuPheSerSerLeuMetLysGlyTyrAsnAlaValLeuSerGlyAsnI 860
2708 TAAACCTTGACAACGCCACCATTACACTCAATTCCGCCTATCGCCACGAT 2757
860 leThrGlyGluGlnSerThrValAsnMetAsnAsnAlaLeuTrpTyrSer 876
2758 GCGGCAGGGGCGCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTC 2807
877 AspGlyAsnSerThrIleGly..... 883
2808 GCGCCGTTCGCGCCGTTCCCTATTATCCGTTACACCGCCAACTTCGGTAG 2857
883 883
2858 AATCCCGTTTCAACACGCTGACGGTAAACGGC.....AAATTGAAC 2898
884 .....ThrLeuLysSerThrGlyGlyArgValGluLeuGly 895
2899 GGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACCGCAGCGA 2948
896 GlyGlyLysAspPheAlaThrLeu..... 903
2949 CAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTACACCTTGGCGGTCA 2998
904 .ArgValLysGluLeuAsnAlaAsnAsnAlaThrPheLeuMetHisThrA 920
2999 ACAATACCGGCAACGAACCTGCAAGCCTCGAACAATTGACGGTAGTGGAA 3048
920 snAsn.....SerGlnAlaAspGlnLeuAsnValThr... 930
3049 GGAAAAGACAACAAACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAA 3098
931 .....AsnLysLeuLeuGlySerAsn...AsnThrValLeu..... 941
3099 CGAACACGTCGATGCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACG 3148
942 .....ValAspPhe..... 944
3149 GCGAGTTCCGCCTGCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAA 3198
945 .....LeuAsnLysProAlaSerGluMetAsnValThr..... 955
3199 CTCGGCAAGGCAGAAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAG 3248
955 955
3249 CCTTGACGCGCTGATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAA 3298
956 .....LeuIleThrAlaProLysGlySerAspGluLysThrPheT 969
3299 GCGTTGCCGAACCGGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATG 3348
969 hrAlaGly.....ThrGlnGlnIleGlyPheSerAsnValThrProVal 983
3349 CAGGCGGAGGAAGAGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTT 3398
984 IleSerThrGlu.....LysThrAspAspAlaThrLysTrpMetLeuTh 998
3399 GGCGAAACAGCGCGAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCC 3448
998 rGlyTyrGlnThrValSerAspAlaGlyAlaSerLysThrAla..... 1012
3449 GCGCCCGCCGCGCCCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAG 3498
1012 1012
3499 CCCCAACCGCAGCGCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAG 3548
1013 .....ThrAspPheMetAlaSerGlyTyrLy 1021
3549 TGAATTTTCCGCCACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTA. 3597
1021 sSerPheLeuThrGluValAsnAsnLeuAsnLysArgMetGlyAspLeuA 1038
3598 .....GACCGCGTATTTGCCGAAGACCGCCGCAACGCCGTT 3633
1038 rgAspThrGlnGlyAspAlaGlyValTrpAlaArgIleMetAsnGlyThr 1054
3634 TGGACAAGCGGCATCCGGG...ACACCAAACACTACCGTTCGCAAGATTT 3680
1055 .GlySerAlaAspGlyGlyTyrSerAspAsnTyrThrHisValGlnIleG 1071
3681 CCGCGCCTACCGCCAACAAACCGACCTGCGCCAAATCGGTATGCAGAAAA 3730
1071 lyAlaAsp.....ArgLys 1075
3731 ACCTCGGCAGCGGGCGCGTCGGCATCCTGTTTTCGCACAACCGGACCGAA 3780
1076 HisGluLeuAspGlyValAspLeu.....PheThrGlyAlaLe 1088
3781 AACACCTTCGACGACGGCATCGGCAACTCGGCACGGCTTGCCCACGGCGC 3830
1088 uLeuThrTyrThrAspSerAsnAlaSerSerHisAla..... 1100
3831 CGTTTTCGGGCAATACGGCATCGACAGGTTCTACATCGGCATCAGCGCGG 3880
1101 ..PheSerGlyLysThrLysSerValGly...Gly..GlyLeuTyrAlaS 1115
3881 GCGCGGGTTTTAGCAGCGGCAGCCTTTCAGACGGCATCGGAGGCAAAATC 3930
1115 erAlaLeuPheAspSerGlyAlaTyrPheAspLeuIleGly..... 1128
3931 CGCCGCCGCGTGCTGCATTACGGCATTCAGGCACGATACCGCGCCGGTTT 3980
1129 .....LysTyrLeuHisHisAspAsnGln.....TyrThrAlaSerPh 1141

```
 3981 CGGCGGATTCGGCATCGAACCGCACATCGGCGCAACGCGCTATTTCGTCC 4030
      |:::::::::|||   :::   :::   :::   :::   |||
 1141 eAlaSerLeuGlyThrLysAspTyrSerSerHisSerTrpTyrAlaGlyA 1158

 4031 AAAAAGCGGATTACCGCTACGAAAACGTCAATATCGCCACCCCCGGCCTT 4080
        ::::::   |||||||||
 1158 laGluValGlyTyrArgTyr............................... 1164

 4081 GCATTCAACCGCTACCGCGCGGGCATTAAGGCAGATTATTCATTCAAACC 4130

 1164 ..... 1164

 4131 GGCGCAACACATTTCC.............ATCACGCCTTATTTGAGCCTGT 4168
            |||:::|||             :::   |||   ::::::|||
 1165 .......HisLeuSerGluGluSerTrpValGluProGlnMetGluLeuV 1179

 4169 CCTATACCGATGCCGCTTCGGGCAAAGTCCGAACACGCGTCAATACCGCC 4218
        |||       ::::::|||||||||
 1179 alTyr...GlySerValSerGlyLys....................... 1186

 4219 GTATTGGCTCAGGATTTCGGCAAAACCCGCAGTGCGGAATGGGGCGTAAA 4268
                                  |||   :::|||
 1187 ..............................SerPheSerTrp........ 1190

 4269 CGCCGAAATCAAAGGTTTCACGCTGTCCCTC 4299
          |||   :::|||::::::|||||||:::
 1191 ....GluAspArgGlyMetAlaLeuSerMet 1199
```

```
seq_name: sp_bacteria:Q9F6T1

seq_documentation_block:
ID   Q9F6T1      PRELIMINARY;      PRT;  1295 AA.
AC   Q9F6T1;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE   SECRETED AUTOTRANSPORTER TOXIN.
GN   SAT.
OS   Escherichia coli.
OC   Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC   Escherichia.
OX   NCBI_TaxID=562;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CFT073;
RX   MEDLINE=20485608; PubMed=11029690;
RA   Guyer D.M., Henderson I.R., Nataro J.P., Mobley H.L.;
RT   "Identification of Sat, an autotransporter toxin produced by
RT   uropathogenic Escherichia coli.";
RL   Mol. Microbiol. 38:53-66(2000).
DR   EMBL; AF289092; AAG30168.1; -.
DR   MEROPS; S06.004; -.
DR   InterPro; IPR000710; IgA_S6.
DR   Pfam; PF02395; IGA1; 1.
DR   PRINTS; PR00921; IGASERPTASE.
SQ   SEQUENCE   1295 AA;  140042 MW;  81122C69273B8FE5 CRC64;


alignment_scores:
            Quality:  636.50              Length:   1585
              Ratio:  0.866                 Gaps:     78
 Percent Similarity:  46.372   Percent Identity: 22.587

alignment_block:
 US-09-303-518D-649 x Q9F6T1   ..

 Align seg 1/1  to: Q9F6T1  from: 1  to: 1295

      37 CGCAAAGCCCCGAAAACCGGCCGCATCCGCTTCTCGCCTGCTTACTTAGC 86
         |||   :::   ||||||:::|||                    :::||
      25 ArgValSerGlyLysThrAsnArg.............LysLeuValAl 36

      87 CATATGCCTGTCGTTCGGCATT.......................... 108
```

```
         |      ||||||:::::::::
      36 aThrMetLeuSerLeuAlaValAlaGlyThrValAsnAlaAlaAsnIleA 53

     109 ..CTTCCCCAAGCCTGGGCGGGACACACTTATTTCGGCATCAACTACCAA 156
           :::   ::::::||||||                              :::
      53 spIleSerAsnValTrpAla.........................Arg 60

     157 TACTATCGCGACTTTGCCGAAAATAAAGGCAAGTTTGCAGTCGGGGCGAA 206
         |||   |||:::|||:::|||||||||   |||      ||||||
      61 AspTyrLeuAspLeuAlaGlnAsnLysGlyIlePheGlnProGlyAlaTh 77

     207 AGATATTGAGGTTTACAACAAAAAAGGGGAGTTGGTCGGCAAATCAATGA 256
          |||:::   :::      |||:::|||:::      :::
      77 rAspValThrIleThrLeuLysAsnGlyAspLysPheSerPheHisAsnL 94

     257 CAAAAGCCCCGATGATTGATTTTTCTGTGGTGTCGCGTAACGGCGTGGCG 306
           :::   |||       |||||||||   ::::::   :::|||:::|||
      94 euSerIlePro......AspPheSerGlyAlaAlaAlaSerGlyAlaAla 108

     307 GCATTGGTGGGCGATCAATATATTGTGAGCGTGGCACATAACGGCGGCTA 356
         :::   :::|||   :::|||   |||:::||||||||||||
     109 ThrAlaIleGlyGlySerTyrSerValThrValAlaHisAsn........ 122

     357 TAACAACGTTGATTTTGGTGCGGAAGGAAGAAATCCCGAT.......... 396
                                    :::||||||:::
     123 ........................LysLysAsnProGlnAlaAlaGluT 131

     397 ........CAACATCGTTTTACTTATAAAATTGTGAAACGGAATAATTAT 438
                    :::   :::|||::::::|||   |||:::|||
     131 hrGlnValTyrAlaGlnSerSerTyrArgValValAspArgArgAsn... 146

     439 AAAGCAGGGACTAAAGGCCATCCTTATGGCGGCGATTATCATATGCCGCG 488
                                       ::::::|||:::::::::   ||
     147 ............................SerAsnAspPheGluIleGlnAr 154

     489 TTTGCATAAATTTGTCACAGAT.........GCAGAACCTGTTGAAATG. 528
         ||||:::|||||||||:::::::         |||   |||:::|||
     154 gLeuAsnLysPheValValGluThrValGlyAlaThrProAlaGluThrA 171

     529 .....ACCAGTTATATGGATGGG......CGGAAATATATCGATCAAAAT 567
              |||:::|||   |||:::      |||      |||      :::
     171 snProThrThrTyrSerAspAlaLeuGluArgTyrGlyIleValThrSer 187

     568 AATTACCCTGACCGTGTT.........CGTATTGGGGCAGGCAGGCAATA 608
         :::          ::::::         |||   |||:::|||
     188 AspGlySerLysLysIleIleGlyPheArgAlaGlySerGlyGlyThrSe 204

     609 TTGGCGATCTGATGAAGATGAGCCCAATAACCGCGAAAGTTCATATCATA 658
          :::                        |||   ||||||:::
     204 rPhe.....................IleAsnGlyGluSerLysIleSerT 214

     659 TTGCAAGTGCGTATTCTTGG...CTCGTTGGTGGCAATACCTTTGCACAA 705
              ||||||||||||      |||:::::::::::   |||
     214 hrAsnSerAlaTyrSerHisAspLeuLeuSerAlaSerLeuPhe...... 228

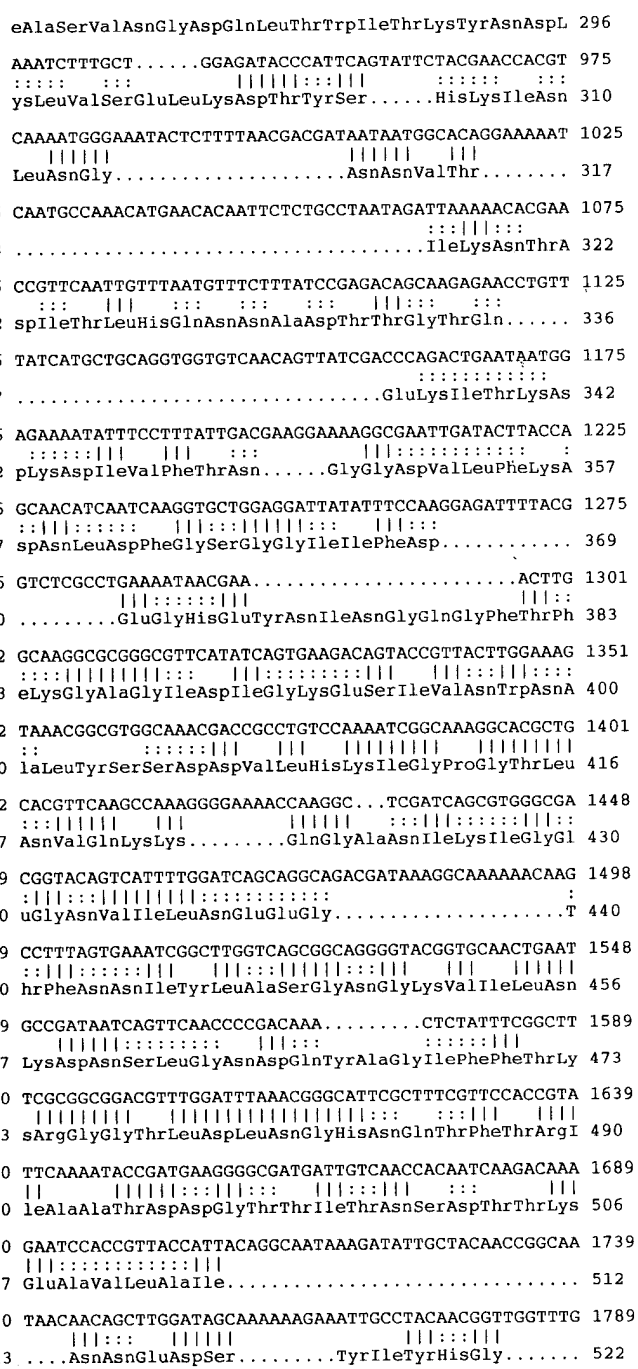
     706 AATGGATCAGGTGGTGGCACAGTCAACTTAGGTAGTGAAAAAATTAAACA 755
                                              |||      :::
     229 ..................................GluValThrGlnTr 233

     756 TAGCCCATATGGTTTTTTACCAACAGGAGGCTCA................ 789
          :::   ||||||::::::      ::::::
     233 pAspSerTyrGlyMetMetIleTyrLysAsnAspLysThrPheArgAsnL 250

     790 ........TTTGGCGACAGTGGCTCACCAATGTTTATCTATGATGCCCAA 831
                 ||||||||||||||||||      ::::::||||||   :::
     250 euGluIlePheGlyAspSerGlySerGlyAlaTyrLeuTyrAspAsnLys 266

     832 AAGCAAAAGTGGTTAATTAATGGGGTATTGCAAACGGGCAACCCCTATAT 881
           :::||||||::::::   |||:::   :::                ||
     267 LeuGluLysTrpValLeuValGlyThrThrHisGly............Il 279

     882 AGGAAAAAGCAATGGCTTCCAGCTGGTTCGTAAAGATTGGTTCTATGATG 931
         |::::::   ||||||   ||||||::                :::   |||:
```


279 eAlaSerValAsnGlyAspGlnLeuThrTrpIleThrLysTyrAsnAspL 296
932 AAATCTTTGCT.....GGAGATACCCATTCAGTATTCTACGAACCACGT 975
296 ysLeuValSerGluLeuLysAspThrTyrSer.....HisLysIleAsn 310
976 CAAAATGGGAAATACTCTTTTAACGACGATAATAATGGCACAGGAAAAAT 1025
311 LeuAsnGly.....AsnAsnValThr..... 317
1026 CAATGCCAAACATGAACACAATTCTCTGCCTAATAGATTAAAAACACGAA 1075
318 .....IleLysAsnThrA 322
1076 CCGTTCAATTGTTTAATGTTTCTTTATCCGAGACAGCAAGAGAACCTGTT 1125
322 spIleThrLeuHisGlnAsnAsnAlaAspThrThrGlyThrGln..... 336
1126 TATCATGCTGCAGGTGGTGTCAACAGTTATCGACCCAGACTGAATAATGG 1175
337 .....GluLysIleThrLysAs 342
1176 AGAAAATATTTCCTTTATTGACGAAGGAAAAGGCGAATTGATACTTACCA 1225
342 pLysAspIleValPheThrAsn.....GlyGlyAspValLeuPheLysA 357
1226 GCAACATCAATCAAGGTGCTGGAGGATTATATTTCCAAGGAGATTTTACG 1275
357 spAsnLeuAspPheGlySerGlyGlyIleIlePheAsp..... 369
1276 GTCTCGCCTGAAAATAACGAA.....ACTTG 1301
370 .....GluGlyHisGluTyrAsnIleAsnGlyGlnGlyPheThrPh 383
1302 GCAAGGCGCGGGCGTTCATATCAGTGAAGACAGTACCGTTACTTGGAAAG 1351
383 eLysGlyAlaGlyIleAspIleGlyLysGluSerIleValAsnTrpAsnA 400
1352 TAAACGGCGTGGCAAACGACCGCCTGTCCAAAATCGGCAAAGGCACGCTG 1401
400 laLeuTyrSerSerAspAspValLeuHisLysIleGlyProGlyThrLeu 416
1402 CACGTTCAAGCCAAAGGGGAAAACCAAGGC...TCGATCAGCGTGGGCGA 1448
417 AsnValGlnLysLys.....GlnGlyAlaAsnIleLysIleGlyGl 430
1449 CGGTACAGTCATTTTGGATCAGCAGGCAGACGATAAAGGCAAAAAACAAG 1498
430 uGlyAsnValIleLeuAsnGluGluGly.....T 440
1499 CCTTTAGTGAAATCGGCTTGGTCAGCGGCAGGGGTACGGTGCAACTGAAT 1548
440 hrPheAsnAsnIleTyrLeuAlaSerGlyAsnGlyLysValIleLeuAsn 456
1549 GCCGATAATCAGTTCAACCCCGACAAA.....CTCTATTTCGGCTT 1589
457 LysAspAsnSerLeuGlyAsnAspGlnTyrAlaGlyIlePhePheThrLy 473
1590 TCGCGGCGGACGTTTGGATTTAAACGGGCATTCGCTTTCGTTCCACCGTA 1639
473 sArgGlyGlyThrLeuAspLeuAsnGlyHisAsnGlnThrPheThrArgI 490
1640 TTCAAAATACCGATGAAGGGGCGATGATTGTCAACCACAATCAAGACAAA 1689
490 leAlaAlaThrAspAspGlyThrThrIleThrAsnSerAspThrThrLys 506
1690 GAATCCACCGTTACCATTACAGGCAATAAAGATATTGCTACAACCGGCAA 1739
507 GluAlaValLeuAlaIle..... 512
1740 TAACAACAGCTTGGATAGCAAAAAAGAAATTGCCTACAACGGTTGGTTTG 1789
513 ....AsnAsnGluAspSer.....TyrIleTyrHisGly..... 522
1790 GCGAGAAAGATACGACCAAAACGAACGGGCGGCTCAACCTTGTTTACCAG 1839
523 .....AsnIleAsnGlyAsnIleLysLeuThrHisAsn 533
1840 CCCGCCGCAGAAGACCGC.....ACCCTGCTGCTTTCCGGCGG 1877
534 IleAsnSerGlnAspLysLysThrAsnAlaLysLeuIleLeuAspGlySe 550
1878 AACAAATTTAAACGGCAACATCACGCAAACAAACGGCAAACTGTTTTTCA 1927
550 rValAsnThrLysAsnAspValGluValSerAsnAlaSerLeuThrMetG 567
1928 GCGGCAGACCAACACCGCACGCCTAC.....AATCATTTA 1962
567 lnGlyHisAlaThrGluHisAlaIlePheArgSerSerAlaAsnHisCys 583
1962 1962
584 SerLeuValPheLeuCysGlyThrAspTrpValThrValLeuLysGluTh 600
1963 .....AACGACCATTGGTCGCAAAAAG 1984
600 rGluSerSerTyrAsnLysLysPheAsnSerAspTyrLysSerAsnAsnG 617
1985 AGGGCATTCCTCGCGGGGAAATCGTGTGGGACAAC...GACTGGATCAAC 2031
617 ln.....GlnThrSerPheAspGlnProAspTrpLysThr 628
2032 CGCACATTTAAAGCGGAAAACTTCCAAATTAAAGGCGGACAGGCGGTGGT 2081
629 GlyValPheLysPheAspThrLeuHisLeuAsnAsnAlaAspPheSerIl 645
2082 TTCCCGCAATGTTGCCAAAGTGAAAGGCGATTGGCATTTGAGCAATCACG 2131
645 eSerArgAsn...AlaAsnValGluGly.....AsnIleSerAlaAsnL 659
2132 CCCAAGCAGTTTTTGGTGTCGCACCGCATCAAAGCCACACAATCTGTACA 2181
659 ysSerAlaIle..... 662
2182 CGTTCGGACTGGACGGGTCTGACAAATTGTGTCGAAAAAACCATTACCGA 2231
663 .....ThrIleGlyAs 666
2232 CGATAAAGTG...ATTGCTTCATTGACTAAGACCGACATCAGCGGCAATG 2278
666 pLysAsnValTyrIleAspAsnLeuAlaGlyLysAsnIleThrAsnAsn. 682
2279 TCGATCTTGCCGATCACGCTCATTTAAATCTCACAGGGCTTGCCACACTC 2328
683 .....GlyPheAspPheLysGlnThrIle 690
2329 AACGGCAATCTTAGTGCAAATGGCGATACACGTTATACAGTCAGCCACAA 2378
691 SerThrAsnLeuSerIle...GlyGluThrLysPheThrGlyGlyIleTh 706
2379 CGCCACCCAAAACGGCAACCTTAGCCTCGTGGGCAATGCCCAAGCAACAT 2428
706 rAla...HisAsnSerGlnIleAlaIleGlyAspGlnAlaValValThrL 722
2429 TTAATCAAGCCACATTAAACGGCAACACATCGGCTTCGGGCAATGCTTCA 2478
722 euAsnGlyAlaThrPheLeuAspAsnThrProIleSer..... 734
2479 TTTAATCTAAGCGACCACGCCGTACAAAACGGCAGTCTGACGCTTTCCGG 2528
735 .....IleAspLysGlyAlaLysValIleAlaGl 744
2529 CAACGCTAAGGCAAACGTAAGCCATTCCGCACTCAACGGTAATGTCTCCC 2578
744 nAsnSerMetPheThrThrLysGlyIleAspIleSerGlyGluLeuThrM 761

```
2579 TAGCC.................GATAAGGCAGTA.........TTCCAT 2601
      ::                   :::|||:::|||         :::|||
 761 etMetGlyIleProGluGlnAsnSerLysThrValThrProGlyLeuHis 777

2602 TTTGAAAGCAGCCGCTTTACCGGACAAATCAGCGGCGGCAAGGATACGGC 2651
     :::   ::::::   |||      ::::::||||||||||:::   :::
 778 TyrAlaAlaAspGlyPhe......ArgLeuSerGlyGlyAsnAlaAsnPh 792

2652 ATTACACTTAAAAGACAGCGAATGGACGCTGCCGTCAGGCACGGAATTAG 2701
      :::      :::   ::::::   |||                    |
 792 eIleAlaArgAsnMetAlaSerValThr.....................G 802

2702 GCAATTTAAACCTTGACAACGCC.......................... 2724
     |||||:::      |||:::|||
 802 lyAsnIleTyrAlaAspAspAlaAlaThrIleThrLeuGlyGlnProGlu 818

2725 ......ACCATTACACTCAATTCCGCCTATCGCCACGATGCGGCAGGGGC 2768
           |||   |||::::::||||||||||:::          |||   ||
 819 ThrGluThrProThrIleSerSerAlaTyrGln.........AlaTrpAl 832

2769 GCAAACCGGCAGTGCGACAGATGCGCCGCGCCGCCGTTCGCGCCGTTCGC 2818
     |:::|||
 832 aGluThr.......................................... 834

2819 GCCGTTCCCTATTA.........................TCCGTTACA 2841
           ||||||                           :::::|||
 835 ........LeuLeuTyrGlyPheAspThrAlaTyrArgGlyAlaIleThr 848

2842 CCGCCAACTTCGGTAGAATCCCGTTTCAACACGCTGACGGTAAACGGCAA 2891
        |||   ::::::   |||      |||::::::
 849 AlaProLysAlaThrValSerMetAsnAsnAlaIle.........TrpHi 862

2892 ATTGAACGGTCAGGGAACATTCCGCTTTATGTCGGAACTCTTCGGCTACC 2941
      ||||||:::|||:::::::::::   ::::::
 862 sLeuAsnSerGlnSerSerIleAsnArgLeuGlu..............T 874

2942 GCAGCGACAAATTGAAGCTGGCGGAAAGTTCCGAAGGCACTTAC...ACC 2988
       :::|||::::::            :::::::::|||   :::   |||
 874 hrLysAspSerMetValArgPheThrGlyAspAsnGlyLysPheThrThr 890

2989 TTGGCGGTCAACAATACCGGCAACGAACCT..................... 3018
     |||:::||||||||||        :::
 891 LeuThrValAsnAsnLeuThrIleAspAspSerAlaPheValLeuArgAl 907

3019 .......GCAAGCCTCGAACAATTGACGGTAGTGGAAGGAAAAGACAACA 3061
           |||:::   :::||||||:::|||             . ||||
 907 aAsnLeuAlaGlnAlaAspGlnLeuValVal..............AsnL 919

3062 AACCGCTGTCCGAAAACCTTAATTTCACCCTGCAAAACGAACACGTCGAT 3111
     ||   ||||||   :::   |||:::   |||            ||||||
 919 ysSerLeuSerGlyLysAsnAsnLeuLeuLeu............ValAsp 931

3112 GCCGGCGCGTGGCGTTACCAACTCATCCGCAAAGACGGCGAGTTCCGCCT 3161
                          :::|||:::|||:::|||
 932 ....................PheIleGluLysAsnGly............ 937

3162 GCATAATCCGGTCAAAGAACAAGAGCTTTCCGACAAACTCGGCAAGGCAG 3211

 937 ..................................................... 937

3212 AAGCCAAAAAACAGGCGGAAAAAGACAACGCGCAAAGCCTTGACGCGCTG 3261
                        ::::::|||:::   :::::::|||   :::
 938 ....................AsnSerAsnGlyLeuAsnIleAspLeuVal 947

3262 ATTGCGGCCGGGCGCGATGCCGTCGAAAAGACAGAAAGCGTTGCCGAACC 3311
      |||              ||||||:::      :::::::::
 948 SerAlaProLysGlyThrAlaValAspValPheLysAlaThr........ 961

3312 GGCCCGGCAGGCAGGCGGGGAAAATGTCGGCATTATGCAGGCGGAGGAAG 3361
      :::|||:::   |||   :::::::|||
 962 .ThrArgSerIleGlyPheSerAspVal..................... 970

3362 AGAAAAAACGGGTGCAGGCGGATAAAGACACCGCCTTGGCGAAACAGCGC 3411
```

```
                                  |||   :::   :::|||:::
 971 ..........................ThrProValIleGluGlnLys 977

3412 GAAGCGGAAACCCGGCCGGCTACCACCGCCTTCCCCCGCGCCCGCCGCGC 3461
     :::          :::   :::   |||   :::                ||
 978 AsnAspThrAspLysAlaThrTrpThrLeuIleGlyTyrLysSerValAl 994

3462 CCGCCGGGATTTGCCGCAACTGCAACCCCAACCGCAGCCCCAACCGCAGC 3511
     |:::   |||      :::
 994 aAsnAlaAspAlaAlaLys............................... 1000

3512 GCGACCTGATCAGCCGTTATGCCAATAGCGGTTTGAGTGAATTTTCCGCC 3561
                  :::        :::::::|||   :::   |||   |||
1001 .....LysAlaThrLeuLeuMetSerGlyGlyTyrLysAlaPheLeuAla 1015

3562 ACGCTCAACAGCGTTTTCGCCGTACAGGACGAATTAGACCGCGTATTTGC 3611
      :::|||::::::
1016 GluValAsnAsnLeu................................... 1020

3612 CGAAGACCGCCGCAACGCCGTTTGGACAAGCGGCATCCGGGACACCAAAC 3661

1020 ..................................................... 1020

3662 ACTACCGTTCGCAAGATTTCCGCGCCTACCGCCAACAAACCGACCTGCGC 3711
                                    :::::::::   |||||||||
1021 ........................AsnLysArgMetGlyAspLeuArg 1028

3712 CAAATCGGTATGCAGAAAAACCTC.................GGCAGCGG 3743
     :::|||:::   :::::::::                    |||:::||
1029 AspIleAsnGlyGluSerGlyAlaTrpAlaArgIleIleSerGlyThrGl 1045

3744 GCGCGTCGGCATCCTGTTTTCGCACAACCGGACC................ 3777
     |   :::|||      ||||||   |||   |||
1045 ySerAlaGlyGlyGlyPheSerAspAsnTyrThrHisValGlnValGlyA 1062

3778 ..GAAAACACCTTCGAC...GACGGCATCGGCAACTCGGCACGGCTTGCC 3822
       :::|||      :::   |||||||:::         :::   ::::::
1062 laAspAsnLysHisGluLeuAspGlyLeuAspLeuPheThrGlyValThr 1078

3823 CACGGCGCCGTTTTCGGGCAATACGGCATCGACAGGTTCTAC........ 3864
               :::   ::::::   |||   |||   |||
1079 MetThrTyrThrAspSerHisAlaGlySerAspAlaPheSerGlyGluTh 1095

3865 .............ATCGGCATCAGCGCGGGCGCGGGTTTTAGCAGCGGCA 3901
                     |||:::   |||:::|||   |||:::||||||:
1095 rLysSerValGlyAlaGlyLeuTyrAlaSerAlaMetPheGluSerGlyA 1112

3902 GCCTTTCAGACGGCATCGGAGGCAAAATCCGCCGCCGCGTGCTGCATTAC 3951
     ::      |||   ||||||               :::   :::|||:::
1112 laTyrIleAspLeuIleGly..............LysTyrValHisHis 1123

3952 GGCATTCAGGCACGATACCGCGCCGGTTTCGGCGGATTCGGCATCGAACC 4001
           :::   :::|||   |||   |||:::|||:::|||
1124 ......AspAsnGluTyrThrAlaThrPheAlaGlyLeuGly........ 1135

4002 GCACATCGGCGCAACGCGCTATTTCGTCCAAAAAGCGGATTACCGCTACG 4051
               ||||||                   ||||||
1136 .............ThrArg.................AspTyr....... 1139

4052 AAAACGTCAATATCGCCACCCCCGGCCTTGCATTCAACCGCTACCGCGCG 4101
                                  :::   :::   :::   |||
1140 ............................SerSerHisSerTrpTyrAla 1146

4102 GGCATTAAGGCAGATTATTCATTCAAACCGGCGCAACACATTTCCATCAC 4151
     |||   ::::::   |||   :::      ::::::         |||
1147 GlyAlaGluValGlyTyrArgTyrHisValThrAspSerAlaTrpIleGl 1163

4152 GCCTTATTTGAGCCTGTCCTATACCGATGCCGCTTCGGGCAAAGTCCGAA 4201
      |||      :::|||   |||      |||:::|||||||||      :
1163 uProGlnAlaGluLeuValTyr...GlyAlaValSerGlyLysGlnPheS 1179

4202 CA............CGCGTCAATACCGCCGTATTGGCTCAGGATTTC... 4236
     ::             :::|||   ::::::      :::||||||
```

```
1179 erTrpLysAspGlnGlyMetAsnLeuThrMetLysAspLysAspPheAsn 1195

4237 ..........................GGCAAAACCCGCAGTGCG.. 4254
                                ||||||:::   |||:::
1196 ProLeuIleGlyArgThrGlyValAspValGlyLysSerPheSerGlyLy 1212

4255 .GAATGGGGCGTAAACGCCGAAATC...AAAGGTTTCACGCTGTCCCTCC 4300
      :::|||   |||:::|||:::         |||:::   ::::::|||
1212 sAspTrpLysValThrAlaArgAlaGlyLeuGlyTyrGlnPheAspLeuP 1229

4301 ACGCT 4305
       |||
1229 heAla 1230
```